

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 28, 2001, 23:43:20 ; Search time 3991.74 Seconds
(without alignments)
2661.333 Million cell updates/sec

Title: US-09-602-597-3
Perfect score: 1516
Sequence: 1 cgtcgcgagcctgggggaaa.....aaaaaaaaaaaaaaaaaa 1516

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 7991742 seqs, 3503743858 residues

Total number of hits satisfying chosen parameters: 15983484

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est25:*
26: gb_est26:*
27: gb_est27:*
28: gb_est28:*
29: gb_est29:*
30: gb_est30:*
31: gb_est31:*
32: gb_est32:*
33: gb_est33:*
34: gb_est34:*
35: gb_est35:*
36: gb_est36:*
37: gb_est37:*
38: gb_est38:*
39: gb_est39:*
40: gb_est40:*
41: em_estba:*
42: em_estfun:*
43: em_esthum1:*

44: em_esthum2:*
45: em_esthum3:*
46: em_esthum4:*
47: em_esthum5:*
48: em_esthum6:*
49: em_esthum7:*
50: em_esthum8:*
51: em_esthum9:*
52: em_esthum10:*
53: em_esthum11:*
54: em_esthum12:*
55: em_esthum13:*
56: em_esthum14:*
57: em_esthum15:*
58: em_esthum16:*
59: em_esthum17:*
60: em_esthum18:*
61: em_esthum19:*
62: em_esthum20:*
63: em_estin1:*
64: em_estin2:*
65: em_estin3:*
66: em_estin4:*
67: em_estov1:*
68: em_estov2:*
69: em_estp1:*
70: em_estp12:*
71: em_estp13:*
72: em_estp14:*
73: em_estp15:*
74: em_estro1:*
75: em_estro2:*
76: em_estro3:*
77: em_estro4:*
78: em_estro5:*
79: em_estro6:*
80: em_estro7:*
81: em_estro8:*
82: em_estro9:*
83: em_estro10:*
84: em_estro11:*
85: em_estro12:*
86: em_estro13:*
87: gb_esta1:*
88: gb_esta2:*
89: gb_esta3:*
90: gb_esta4:*
91: gb_esta5:*
92: gb_esta6:*
93: gb_esta7:*
94: gb_esta8:*
95: gb_esta9:*
96: gb_esta10:*
97: gb_esta11:*
98: gb_esta12:*
99: gb_esta13:*
100: gb_esta14:*
101: gb_esta15:*
102: gb_esta16:*
103: gb_esta17:*
104: gb_esta18:*
105: gb_esta19:*
106: gb_esta20:*
107: gb_esta21:*
108: gb_esta22:*
109: gb_esta23:*
110: gb_esta24:*
111: gb_esta25:*
112: em_esthum21:*
113: em_esthum22:*
114: em_esthum23:*
115: em_estom1:*
116: em_estom2:*

117: em_estp16:*
118: em_estp17:*
119: em_estp18:*
120: em_estp19:*
121: em_estp20:*
122: em_estp21:*
123: em_estp22:*
124: em_estp23:*
125: em_estp24:*
126: em_estp25:*
127: em_estp26:*
128: em_estp27:*
129: em_estp28:*
130: em_estp29:*
131: em_estp30:*
132: em_estp31:*
133: em_estp32:*
134: em_estp33:*
135: em_estp34:*
136: em_estp35:*
137: em_estp36:*
138: em_estp37:*
139: em_estp38:*
140: em_estp39:*
141: em_estp40:*
142: em_estp41:*
143: em_estp42:*
144: em_estp43:*
145: em_estp44:*
146: em_estp45:*
147: em_estp46:*
148: em_estp47:*
149: em_estp48:*
150: em_estp49:*
151: em_estp50:*
152: em_estp51:*
153: em_estp52:*
154: em_estp53:*
155: em_estp54:*
156: em_estp55:*
157: em_estp56:*
158: em_estp57:*
159: em_estp58:*
160: em_estp59:*
161: em_estp60:*
162: em_estp61:*
163: em_estp62:*
164: em_estp63:*
165: em_estp64:*
166: em_estp65:*
167: em_estp66:*
168: em_estp67:*
169: em_estp68:*
170: em_estp69:*
171: em_estp70:*
172: em_estp71:*
173: em_estp72:*
174: em_estp73:*
175: em_estp74:*
176: em_estp75:*
177: em_estp76:*
178: em_estp77:*
179: em_estp78:*
180: em_estp79:*
181: em_estp80:*
182: em_estp81:*
183: em_estp82:*
184: em_estp83:*
185: em_estp84:*
186: em_estp85:*
187: em_estp86:*
188: em_estp87:*
189: em_estp88:*

190: gb_gss25:*
191: gb_gss26:*
192: gb_gss27:*
193: gb_gss28:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	642.8	42.4	954	109	BE563252	BE563252 601335555
2	602.8	39.8	699	109	BE514802	BE514802 601335734
3	587.8	38.8	839	135	BE742339	BE742339 601575575
4	541.8	35.7	718	38	AV708191	AV708191 AV708191
5	534.8	35.3	539	25	AI801867	AI801867 kx28e03.x
6	492.4	32.5	494	111	BE572806	BE572806 7d24d08.x
7	483	31.9	495	26	AI870611	AI870611 v175e02.x
8	479	31.6	479	27	AI991347	AI991347 wu42h12.x
9	473	31.2	489	14	AA959543	AA959543 ou98a07.s
10	467.4	30.8	469	91	AA471404	AA471404 xw62901.x
11	465	30.7	466	23	AI683914	AI683914 tw54f12.x
12	459	30.3	459	26	AI928415	AI928415 w067e11.x
13	450.6	29.7	457	13	AA909585	AA909585 o118f12.s
14	440.4	29.1	442	23	AI683813	AI683813 kx77d10.x
15	424.4	28.0	426	90	AA468787	AA468787 hc28a08.x
16	418.8	27.6	457	13	AA918424	AA918424 o170h01.s
17	406	26.8	406	19	AI359905	AI359905 qy35b03.x
18	404	26.6	412	90	AA469186	AA469186 hc78q08.x
19	401.4	26.5	411	19	AI358220	AI358220 gw14a04.x
20	393	25.9	420	15	AI050920	AI050920 gw32h11.s
21	384	25.3	441	141	H30310	H30310 y077h04.t1
22	374.6	24.7	387	15	AI028514	AI028514 ow44b04.x
23	360.2	23.8	365	136	BE857931	BE857931 7f72b08.x
24	354.2	23.4	388	143	RI6135	RI6135 ya51h07.s2
25	350.4	23.1	551	23	AI688299	AI688299 w087h10.x
26	348	23.0	387	143	RI6134	RI6134 ya51h06.s2
27	341	22.5	374	24	AI739570	AI739570 w135b07.x
28	328	21.6	329	27	AI991652	AI991652 wf15c08.x
29	320.6	21.1	371	38	AV708918	AV708918 AV708918
30	319.8	21.1	323	91	AM516060	AM516060 xl60a01.x
31	318.4	21.0	328	5	AA302987	AA302987 EST113187
32	310	20.4	310	13	AA918497	AA918497 o173h04.s
33	305.8	20.2	461	143	RI6034	RI6034 ya51h08.t2
34	294.6	19.4	311	140	D25700	D25700 H0MG504.066
35	272	17.9	272	147	Z38289	Z38289 HSC05E04.2 n
36	270.8	17.9	281	147	Z42004	Z42004 HSC05E04.1 n
37	266.8	17.6	271	5	AA298803	AA298803 EST114404
38	266	17.5	306	111	BE702436	BE702436 RC5+AN106
39	261.2	17.2	322	111	BE702431	BE702431 RC5+AN106
40	260.8	17.2	317	91	AA480865	AA480865 33897.MAR
41	255.2	16.8	475	143	RI6035	RI6035 ya51h07.t2
42	229.2	15.1	939	135	BE735910	BE735910 601335285
43	215.4	14.2	250	5	AA302929	AA302929 EST113185
44	187.4	12.4	245	89	AM345513	AM345513 15033.MAR
45	170.4	11.2	172	140	F01529	F01529 HSC036.02 n

ALIGNMENTS

RESULT 1
BE563252 954 bp mRNA EST. 15-AUG-2000
LOCUS 601335555F1 NIH_MGC_39 Homo sapiens cDNA clone IMAGE:3689653 5',
DEFINITION BE563252 mRNA sequence.
ACCESSION BE563252
VERSION BE563252.1 GI:9807064
KEYWORDS EST.
SOURCE human.

QY	1116	aaacacggtttcaaccacccctcctctgagccaaactgtataaagvtggttattgctg	1175
Db	181	AAACACCGCTTTCAACCAACCTCTCCCTGGAGCCAAACCTGTAAAAAGTGGTTGATGCTG	240
QY	1176	acagatagctctccctccctccgcatttaagacataccggttaccgaagcaatcag	1235
Db	241	ACAGATAGCTCTCCCTCCCTCCCTGCAATTTAGACATACAGTTACTGAAAGCAAAATCAGTT	300
QY	1236	taagtattcttcagtgctgtaaaagccctgcaggttccctcccttccccaagcttc	1295
Db	301	TAAAGTATTTCGCAATGCTGAAAGCCGTGCAGGTTTCCTCCCTTCCCAAGCTTC	360
QY	1236	tctgtaatactccctctgaggcgaagctaacaatcggtgctccctccgaacctgtactagg	1355
Db	361	TCTGTAAATACCTCCCTTTGGCGAAGCTAACTACGTGGTGCTTCCGACCTTCTGCTAGG	420
QY	1356	cacatggagcgcgaagaggagggaagcaagacctgtccctggcgagtgatgtcatggtt	1415
Db	421	CACATGGAGCCGCAAGGAGGGAAGCAAGGCCCTTGGCGAGTGTTCATGTGCTT	480
QY	1416	ggtgtgacctgttatttttttaataaataaagaatgagagaattaaaaaaa	1475
Db	481	GGTGTGACTGCTGTTATTTTATTTATATAAATAAAGATGAGAGAAATTTATTTGGAAAAA	540
QY	1476	aaaaaaaaaaaaa	1489
Db	541	AAAAAAAAAAAAA	554
RESULT	5		
LOCUS	AI801867		
DEFINITION	tx8be03.x1 NCI_CGAP_Lu24 Homo sapiens CDNA IMAGE:2270908 3',		
VERSION	AI801867		
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Mammalia; Euthelia; Primates; Catarrhini; Homnidae; Homo.		
AUTHORS	1 (bases 1 to 539)		
TITLE	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.		
JOURNAL	National Cancer Institute, Cancer Genome Anatomy Project (CGAP),		
COMMENT	Tumor Gene Index		
	Unpublished (1997)		
	Contact: Robert Strausberg, Ph.D.		
	Tel: (301) 496-1550		
	Email: Robert.Strausberg@nih.gov		
	Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.		
	Emmert-Buck, M.D., Ph.D.		
	CDNA Library Preparation: M. Bento Soares, Ph.D.		
	CDNA Library Arrayed by: Greg Lennon, Ph.D.		
	DNA Sequencing by: Washington University Genome Sequencing Center		
	Clone distribution: NCI-CGAP clone distribution information can be		
	found through the I.M.A.G.E. Consortium/LNL at:		
	www-bio.llnl.gov/bdip/image/image.html		
	Insert Length: 673 Std Error: 0.00		
	Seq primer: -400P from Gdbco		
	High quality sequence stop: 470.		
FEATURES	location/Qualifiers		
SOURCE	1. 539		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone="IMAGE:2270908"		
	/clone_lib="NCI_CGAP_Lu24"		
	/tissue_type="carcinoid"		
	/lab_host="DH10B"		
	/note="Organ: Lung; Vector: pUT73D-Pac (Pharmacia) with a		
	modified polylinker; Plasmid DNA from the normalized		
	library NCI_CGAP_Lu5 was prepared, and ss circles were		
	made in vitro. Following HAP purification, this DNA was		

used as tracer in a subtractive hybridization reaction.
The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (cloneds 1414920-1417991 and 1520904-1522439). Subtraction by Bento Soares and M. Fatima Bernaldo.

BASE COUNT 127 a 158 c 124 g 129 t 1 others
ORIGIN

Query Match 35.3%; Score 534.8; DB 25; Length 539;
Best Local Similarity 99.4%; Pred. No. 1.4e-101;
Matches 536; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 875 catccccccttcaagtggaagatgacgggggagcttgagctgtgtctgtctgtctgt 934
DB 1 CATCCCCCTTATTCAGTGGAGATGACGGGGATGTGAGGCTGTCTGTGCTTGT 60
QY 935 tagaggaactcaagtcgaagactgaggccaccccttccaccagactaaatgactaa 994
DB 61 TAGAGGACTTCAGGCTCCAAAGACTGGGGCCACCTTCTCACAGCACTAAATGACTAA 120
QY 995 caaaggaactcgaagctgagcccccagaccgcccgaatgataagcctaaagaagaacagt 1054
DB 121 CAAGGACTCCAGACTGACGCCGCCAGACCCGCGATATTAAGCCTTAAGAAAGCAACAGCT 180
QY 1055 agcaactagctgtgtgtccagagagactgagcaagctgtgtgaaacacactcttccct 1114
DB 181 AGCACTTAGCTTTGTGTCCAGAGAGCTGAGCAAGCTGTGGAACCACTCCTCCTTCT 240
QY 1115 taacaacccgttcaacaacactctccctgagagcaacactgtgaaagtgggtgtgtgt 1174
DB 241 TAAACACCGTTCACCAACCACTCCTCCGAGCCCAACTGTAAAGAGGGTGTGATTTCT 300
QY 1175 gacagcaatggtctccctccctccctcctcctcctcctcctcctcctcctcctcct 1234
DB 301 GACAGCACTGAGCTGCTCCCTCCCTGCTTTCACACATACAGTTACTGAAAGCAATCACT 360
QY 1235 ttaagtgatcttcagtgagtgtaaaagcctgtccaggttctcctcctcctcctcctcct 1294
DB 361 TTAAGTGTATTCAGTGTCTGAAAGCCTGTCCAGGTTTCTTCCCTTCCCAAGCCTCT 420
QY 1295 ctctgttaactcccttggggcgaagtaacacagtgagctcctccagactgtgtgactag 1354
DB 421 CTCTGTATTAATCTCCCTTGGGGCGAAGCTAATCATGTGCTCCCTCCGACCTTGTACTAG 480
QY 1355 gcaatggaagcgaagaagaggaaggaagcagagcctgtcctgagacttgcattgtg 1413
DB 481 GCAATGGGAGCGCANAGAGGAGGAGGAGCAAGGCTTGTGCTGGGAGTGTGATGTGG 539

RESULT 6 BE672806 494 bp mRNA EST 08-SEP-2000
LOCUS BE672806/c 7024a08.x1 NCI_CGAP_Pt28 Homo sapiens cDNA clone IMAGE:3248150 3'
DEFINITION Similar to contains Alu repetitive element.; mRNA sequence.
ACCESSION BE672806
VERSION BE672806.1 GI:10033347
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 494)
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncigap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.
Zimmer-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.

CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL, send email to:
info@image.lnl.gov
Seq primer: -400p from Glibco
High quality sequence stop: 463.
Location/Qualifiers

FEATURES

source

1..494
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3248150"
/clone_id="NCI_CGAP_Pt28"
/sex="male"
/dex_stage="adult"
/lab_host="DH10B"

/note="Organ: prostate; Vector: pT73D-Pac (Pharmacia)
with a modified polylinker. Plasmid DNA from the
normalized library NCI_CGAP_Pt22 was prepared, and ss
circles were made in vitro. Following HAP purification,
this DNA was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from a pool
of 5,000 clones made from the same library (cloneds
983608-986759, 1101192-1101959, and 1217928-1220615).
Subtraction by Bento Soares and M. Fatima Bernaldo.

BASE COUNT 128 a 107 c 130 g 129 t
ORIGIN

Query Match 32.5%; Score 492.4; DB 11; Length 494;
Best Local Similarity 99.8%; Pred. No. 9.3e-93;
Matches 493; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 971 tctaacgacgaactaatcactaaagaagactcgaagctcgaagcccgccgcta 1030
DB 494 TCTCACGAGCTTAATCTACTTAACAGAGACTCCAGACTCAGCCGACCGCCTA 435
QY 1031 gataagcctaaagaacacacgtagacacttagcttctgtccagagagctgagcaag 1090
DB 434 GTATTAAGCTTAACAAGCAACAGTAGACCTTAAGCTTTGTGTCCAGAGAGACTGAGCAG 375
QY 1091 ctgtgaaacacactctcctcctttaaaccacggttcaaccaactctcctgtgagccaa 1150
DB 374 CTGTGAACACGACTCTCTCTTTAAACACGTTTCAACCAACTCTCTCTGAGGCCAA 315
QY 1151 cctgtaaaagtgtgtatgtgtcgaagcagtggtctcctcctgtgattcagacta 1210
DB 314 CTTGTAAGAGTGGGTGATGTGCTGACGACATGGCTTCTCCCTGATTTCAAGACTA 255
QY 1211 ccagttactgaagcaaacacagtttaagtgattctcagtgctgaaagctgtccag 1270
DB 254 CCAGTTACTGAAGCAATCATGTTTAAGTATTTCTCAGTGTGAAAGCTGTGCCAG 195
QY 1271 ttctcctccttcccaagcctctctgttaatacctccttggggcgaagctaacatcg 1330
DB 194 TTTCTCTTCCCTTTCACAGCCTCTCTGTATTAATCTCTCTTGGGCGAAGCTAACATCGG 135
QY 1331 tgcctcccccagcctgtgactaagcacatgagcagcaagaaggaaggaaggaagcc 1390
DB 134 TGCTCCCGGAGCTTGTGCTGACTAGGACATGGAGCGCAAGAGGAGGAGGAGCAAGGCC 75
QY 1391 ttgcctgagcagtgctcagtggtgtgtgtgactgttcttctttaaataaataa 1450
DB 74 TTGCTGGCGAGTGTGATGATGTTGGTGGTGAGTGTATTTTATTAATAAATAA 15
QY 1451 agatgagagaact 1464
DB 14 AGATGAGAGAAATT 1

RESULT 7 A1870611/c 495 bp mRNA EST 07-MAR-2000
LOCUS A1870611

[illegible]

Tel: (301) 496-1550
 Email: Robert.Strausberg@nih.gov
 Life Technologies catalog #: 11548-013
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/BLNL at:
 www.bio.lnl.gov/bdrr/image/image.html
 Seq primer: -40UP from Glibco
 High quality sequence stop: 424.

FEATURES

source

1. 469
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:2832624"
 /clone_lib="NCI-CGAP_Pan1"
 /tissue_type="adenocarcinoma"
 /lab_host="DH10B"
 /note="Organ: pancreas; Vector: PCMV-SPORT6; Site:1: Salt;
 Site:2: NotI; Cloned unidirectionally. Primer: Oligo dT.
 Average insert size 1.72 kb. Life Technologies catalog #:
 11548-013"

BASE COUNT 124 a 104 c 121 g 120 t
 ORIGIN

Query Match 30.8%; Score 467.4; DB 91; Length 469;
 Best Local Similarity 99.8%; Pred. No. 1.5e-87;

Matches 468; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 997 agactccaaactgcagcccgccgagctagataagctaaacacacacagtag 1056
 |||||
 DB 469 AGACTCCAACTGCAGCCCGCCGCTGATAGTATAGCTTACACACACACAGTAG 410
 QY 1057 caacctagctctgttccagagagagctgagcaagctggtgaaacacactcctcctta 1116
 |||||
 DB 409 CACCTAGCTCTGTTCAGAGAGAGCTGAGCAGAGCTGAGTGAACCACTCTCTCTCTTA 350
 QY 1117 aaacccgtttcaacccaactctccctgaggccaactgtataaagtggtgtgctgta 1176
 |||||
 DB 349 AACACCGTTTCAACCAACCTCTCCCTGAGACCAACTGTAAAGTGAGTGTGCTGA 290
 QY 1177 caagcagctctccctccctgcattccagacatacagttactgaaagcaaatcagttt 1236
 |||||
 DB 289 CAGCATGCTCTTCCCTCCCTGCACTTTCAGACATACCAAGTACTGAAAGCAATCACTTTT 230
 QY 1237 aagtgatctcagtgctgaaagcctgtccaggttcccttccctcccaagcctctct 1296
 |||||
 DB 229 AAGTATTTCTCAGTCTGAAAAAGCCTGTCCAGTTCCTTCCCTTCCCAAGCCTCTCT 170
 QY 1297 ctgtaatactccctcttggggcgaagcacaatcggtgctcccccagacttgctgactaagc 1356
 |||||
 DB 169 CTGTATTAATCTCCCTTGGGCGAAGCTTAACATCGTCCCTCCGACCTTGCTGACTAGGGC 110
 QY 1357 acatggagcgaaggaaggaaggaagcctgctgctgagagtttcatggtgtg 1416
 |||||
 DB 109 ACATGGAGGCAAGAGAGGAGGAGCAAGGCTTGCTGCGAGATTGTGCATGTGGTGG 50
 QY 1417 gtggtagctgtttatttttttaataaataaagaatgagagaata 1465
 |||||
 DB 49 GTGGTAGCTGTATTATTTTATTAATAAATAAAGATGAGAGAAATTA 1

RESULT 11
 A1683914 466 bp mRNA EST 15-DEC-1999
 LOCUS A1683914/C
 DEFINITION tw54f12.x1 NCI-CGAP U1 Homo sapiens cDNA clone IMAGE:2263535 3'
 similar to contains Alu repetitive element; mRNA sequence.
 ACCESSION A1683914
 VERSION A1683914.1 GI:4894096
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE 1 (bases 1 to 466)
 AUTHORS NCI-CGAP
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert.Strausberg@nih.gov
 Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
 Emmert-Buck, M.D., Ph.D.
 CDNA Library Preparation: Life Technologies, Inc.
 DNA Sequencing by: Greg Lennon, Ph.D.
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/BLNL at:
 www.bio.lnl.gov/bdrr/image/image.html
 Insert Length: 1587 Std Error: 0.00
 Seq primer: -40UP from Glibco
 High quality sequence stop: 405.

FEATURES

source

1. 466
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:2263535"
 /clone_lib="NCI-CGAP_U1"
 /tissue_type="well-differentiated endometrial
 adenocarcinoma, 7 pooled tumors"
 /lab_host="DH10B"
 /note="Organ: uterus; Vector: PCMV-SPORT6; Site:1: Salt;
 Site:2: NotI; Cloned unidirectionally. Primer: Oligo dT.
 Average insert size 1.75 kb. Life Technologies catalog #:
 11538-014"

BASE COUNT 121 a 100 c 116 g 128 t 1 others
 ORIGIN

Query Match 30.7%; Score 465; DB 23; Length 466;
 Best Local Similarity 99.8%; Pred. No. 4.8e-87;

Matches 465; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1012 caagcccaagacccgagctagataagccttaagcaagcaacagctagctttgt 1071
 |||||
 DB 466 CAGCNCAGACCGCCGCTATATAGCTATACAGCAACAGTACCTTAGCTTTGT 407
 QY 1072 tccaagagagctgagcaagctgtgaaacacactcctccttccctaaacacggttcaacc 1131
 |||||
 DB 406 TCCAGGAGACTGAGCAAGCTGTGAACACACTCTCTCTTTAAACACGTTTCAAC 347
 QY 1132 aacctcctctgagccaactgtaaaagtgtgtgattgtctgacgaatggtctccc 1191
 |||||
 DB 346 AACCTCTCTCTGAGGCAACCTGTAAAGTGAGTGTGACGACATGCTCTTCCC 287
 QY 1192 tcccttcatttgagatcccaatcactgtaaaagaacatgattgaagtattccagt 1251
 |||||
 DB 286 TCCCTGCAATTCAGACATACCAAGTACTGAAACCAATCAGTTTAAAGTATTTTCAG 227
 QY 1252 gctgaaaagcctgctcaggttctccttccctcccaagcctcctcctgtaatactccct 1311
 |||||
 DB 226 GCTGAAAAGCTGTCCAGGTTTCTTCCCTTCCCAAGCCTCTCTGTATATCTCCCT 167
 QY 1312 tggggaagcctaactcagtgctcccgagccttgctgactaggaacatgaggagcaag 1371
 |||||
 DB 166 TGGGCAAGAGCTTAACATCGGCTCCCGACCTTGCTGACTAGGACATGGGAGCGCAAG 107
 QY 1372 gaggagaggaaggaagccttgctgctgagagtgatgtgtgtgtgtgtgtgtgtgtgt 1431
 |||||
 DB 106 GAGGGAGGAGCAAGGCTTGGCTTGGCGAGTGTGATGTTGGTGGTGGTGGTGGTGGT 47
 QY 1432 ttttttataaataaagaatgagagaatgagagaatgagagaatgagagaatgagagaat 1477
 |||||
 DB 46 TTTTATTAATAAATAAAGATGAGAGAAATTAATAAATAAATAAATAAATAAATAAATAA 1

RESULT 12
AI928415/c 459 bp mRNA EST 08-MAR-2000
LOCUS w067e11.x1 NCI_CGAP_Pr22 Homo sapiens CDNA clone IMAGE:2460428 3'
DEFINITION similar to contains Alu repetitive element; mRNA sequence.
ACCESSION AI928415.1 GI:5664379
VERSION AI928415
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 459)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
TITLE Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LINL at:
www-bio.llnl.gov/bdnp/image/image.html
Insert Length: 1504 Std Error: 0.00
Seq primer: -400p from Glbco.
Location/Qualifiers
source
1. .457
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2460428"
/clone_id="NCI_CGAP_Pr22"
/sex="male"
/tissue_type="normal prostate"
/lab_host="DH10B"
/note="Organ: prostate; Vector: p773D-Pac (Pharmacia)
with a modified polylinker. 1st strand CDNA was prepared
from normal prostate bulk tissue, and was then primed with
a Not I - oligo(dT) primer. Double-stranded CDNA was
ligated to Eco RI adaptors (Pharmacia), digested with Not
I and cloned into the Not I and Eco RI sites of the
modified p773 vector. Library is normalized, and was
constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 121 a 102 c 122 g 114 t
ORIGIN
Query Match 30.3%; Score 459; DB 26; Length 459;
Best Local Similarity 100.0%; Pred. NO. 8.5e-86;
Matches 459; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1001 ctccagacctgcaagcccgccgtagtataagcctaacaagcaacagtagcacc 1060
|||||
DB 459 CTCGAGACCTGACGCCGACGACCGCCTAGTATTAAGCTTAACAAGCAAGTAGAGCACC 400
QY 1061 ttagctcttgcagagagagctgagcaagctggtgaacaacctctcctttaaaca 1120
|||||
DB 399 TTACTCTTTGTTCCAGGAGAGAGCTAGCAAGCTGTGGAACACACTCTCTCTTAACA 340
QY 1121 ccggttcaacaacctctccctgagagccaccttaaaaatggtgattgctgagagc 1180
|||||
DB 339 CCGTTTCAACCAACCTCTCTCTGAGGCCAAGCTGAAAGTGGGTGATTGCTGACAGC 280
QY 1181 atggcttccctccctgattcagacacatcagttactgaaagcaaatcagtttaagt 1240
|||||
DB 279 ATGGCTTCCCTCCCTGATTTACAGACTATACAGTATGAAAGCAATAGTTTAAGT 220
QY 1241 gattctcagtgctgaaagcctgccaagttccttcccttccccaagcctctctgt 1300

RESULT 13
AA909585/c 457 bp mRNA EST 23-JUN-1998
LOCUS o118f12.s1 Soares_NFL_T_GBC_S1 Homo sapiens CDNA clone
DEFINITION IMAGE:1523855 3', mRNA sequence.
ACCESSION AA909585
VERSION AA909585.1 GI:3048990
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 457)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
TITLE Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
This clone is available royalty-free through LINL; contact the
IMAGE Consortium (infoimage.llnl.gov) for further information.
Insert Length: 691 Std Error: 0.00
Seq primer: -40m13 fwd. RT from Amersham
High quality sequence stop: 434.
Location/Qualifiers
source
1. .457
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1523855"
/clone_id="Soares_NFL_T_GBC_S1"
/lab_host="DH10B"
/note="Organ: pooled; Vector: p773D-Pac (Pharmacia) with
a modified polylinker; Site_1: Not I; Site_2: Eco RI;
Equal amounts of plasmid DNA from three normalized
libraries (fetal lung NBHL19W, testis NHT, and B-cell
NCL-GARP GCB1) were mixed, and as circles were
in vitro. Following HAP purification, this DNA was used as
tracer in a subtractive hybridization reaction. The driver
was PCR-amplified cDNAs from pools of 5,000 clones made
from the same 3 libraries. The pools consisted of
I.M.A.G.E. clones 297480-302087, 682632-687239,
726408-728711, and 729096-731399. Subtraction by Bento
Soares and M. Fatima Bonaldo."

BASE COUNT 121 a 104 c 117 g 115 t
ORIGIN
Query Match 29.7%; Score 450.6; DB 13; Length 457;
Best Local Similarity 99.1%; Pred. NO. 4.8e-84;
Matches 453; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1011 gcaagcccaagaccgcccgtatataagcctaacaagcaacagtagcacccttcttg 1070
|||||
DB 457 GCAGCCCAAGACCGCCGCTAGTATTAAGCTTAACAAGCAAGTAGAGCACCCTTAAGTCTTTG 398
QY 1071 ttcaagagagctgagcaagctggtgaacaacctctcctttaaaccacggttcaac 1130

Db	397	TTCCAGGAGAGGTGAGCAAGCTGGGGAACCACTCTCTTCCTTTAAACACCGTTTCAC	338
QY	1131	caacctctccctcgtgagccaacctgtgtaaaagltggttgaattgctcgtacagcaatgltctcc	1190
Db	337	CAACCTCTCCCTGGAGCCACACTGTAAAAAGGGTTATTGTGCAGCAGCANTGGCTTCC	278
QY	1131	ctccctgcatttcagacataccagttacttgaagaacaatcagtttaagtattctcag	1250
Db	277	CTCCCTGCATTTCAAGACATACCAGTTACTGAAAGCAANTCAGTTTAAAGTATTTTCAG	218
QY	1251	tgcgtgaaaagcctgtcgaagtttcccttcccttcccaagcctctctctgtataatccct	1310
Db	217	TGCTGAAAAAGCCTGTCACAGTTTCCTTCCCTTCCCAAGCCTCTCTCTGTAATCTCGCT	158
QY	1311	ttaggcgcgaagctaacatcgtgtgctctcccgaccttgcctgactaagacacatvgygaacgaa	1370
Db	157	TTGGGCGAAGCTAACATCTGGTGCTCCCGCCGACCTTGTCTGACATAGGCAANTGGAGCGCAA	98
QY	1371	ggaggggaaggaggaagccttgcctgycaggttgtatlgtygttgytgygactgtttt	1430
Db	97	GGAGGGAGGGAAGCAAGGCTTGGCTCGGCGAGTTGTCAATGTGGTGGTGAGTGTGTTT	38
QY	1431	atttttttaataaaataaagatgtaggaagaaattaa	1467
Db	37	ATTTTTTTTAAATAAATAAAGATGGGCGGAATTAAA	1

RESULT 14	LOCUS	DEFINITION	ACCESION
A1683813/C	A1683813	442 bp mRNA EST	16-DEC-1999
	tx77d10.x1	NCI_CGAP_Ucl Homo sapiens cDNA clone IMAGE:2275603	3'
		similar to contains Alu repetitive element; mRNA sequence.	
	A1683813		

ORGANISM

REFERENCE

JOURNAL
COMMENT

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.lnl.gov/dbp/rimage/image.html
Insert Length: 889 Std Error: 0.00
Seq primer: -400P from Gibco
High quality sequence stop: 405.

FEATURES

SOURCE

BASE COUNT

ORIGIN

Query Match	29.1%;	Score 440.4;	DB 23;	Length 442;
Best Local Similarity	99.8%;	Pred. No. 6.4e-82;		
Matches 441;	Conservative 0;	Mismatches 1;	Indels 0;	Gaps 0;

QY	1050	caagtaacacctaagtccttggttccagagagctgaagcaagcttggtgaacacatctcc	1109
Db	442	CAGTACACACTTATGCTTTGTTGTCAGGAGAGCTGAGCAAGTGGTGAACCACTTCT	383
QY	1110	tcctttaaacacccgtttcaaccaacacctctccctgagagccaacctgttaaagttggtlga	1169
Db	382	TCTTTTAAACACCGTTTCAACCAACCTTCCCTGGAGCCAACTGTATAAAAGTGGTTGA	323
QY	1170	ttggtgacagcattggtcttccctccctgcatcttcagacatcacgtttacgtgaagcaaat	1229
Db	322	TTGTGTGACACATGAGTCCTTCCCTCCCTGGATTTTCAGACATACCAATTTACTGAAAGCAAT	263
QY	1230	caagtttaagtgattcttcagctgctgaagaagctgtccaggttcttcccttcccaag	1289
Db	262	CAGTTTAAATGATTTCTCAGTGTCTAANAAGCTGTCCAGGTTTCTTCCCTTCCCAAG	203
QY	1290	cctctctcgttaactcccttggcggaagctaaacatcgtctgcctcccgacactgctg	1349
Db	202	CCCTCTCTGTATACCTCCCTTGGGCGAAGCTAACATGAGTCCCTCCCAACCTTGCTG	143
QY	1350	actcaggaactgtagaagcaagaaggaaggaaggaagaagccttgcctctggaagtgtcat	1409
Db	142	ACTTAGGACACTGGAGACGCAAGAGGAGGAGGAGCAAGGCTCTGCTGCGAGTTGTAT	83
QY	1410	gtgggttggtgactggttttatcttctttaaataaataaagaatcgagagaaatlaaaa	1469
Db	82	GTGGTGTGTTGAGCTGTTTATTTTATTTTAAATAAATAAAGATGAGAGAAATTAATAA	23
QY	1470	aaaaaaaaaaaaaaaaaaaaa 1491	
Db	22	AAAAAAAAAAAAAAAAAAAAA 1	

RESULT 15

LOCUS	426 bp	mRNA	EST	24-FEB-2000
DEFINITION	h228a08.x1 Soares_NFL_T_GBC_S1 Homo sapiens	CDNA	clone	
IMAGE:2910806	3'	mRNA	sequence.	

SOURCE ORGANISM

REFERENCE

TITLE

JOURNAL
COMMENT

FEATURES

```

'organism="Homo sapiens"
'ab_xref="taxon:9606"
'clone_1b="IMAGE:2910806"
'clone_1lb="Soares_NFL_T_GBC_S1"
'ab_host="DH10B"
'note="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with
a modified polylinker; Site_1: Not I; Site_2: Eco RI;

```

Equal amounts of plasmid DNA from three normalized libraries (fetal lung NDHL19W, testis NHT, and B-cell NCI-GAP-GCB1) were mixed, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 3 libraries. The pools consisted of I.M.A.G.E. clones 297480-302087, 682632-687239, 726408-728711, and 729096-731399. Subtraction by Bento Soares and M. Fatima Bernaldo.

BASE COUNT 117 a 94 c 104 g 111 t
ORIGIN

Query Match 28.0%; Score 424.4; DB 90; Length 426;
Best Local Similarity 99.8%; Pred. No. 1.4e-78;
Matches 425; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1043 caagcaacagctagacacttcttcttccagagagctlgagcaagctggtgaacca 1102
|||||
DB 426 CAAGCAACAGCTAGACACTTCTTGTTCAGAGAGCTGAGCAAGCTGGTGAACCA 367
OY 1103 ctctccctcttaaacacgcttcaacacactctccctgagccaaactgtaaaagt 1162
|||||
DB 366 CTCTCCCTCTTAAACACCGTTTCAACCAACCTCTCCATGAGCCCAACCTGTAAAGT 307
OY 1163 ggggtgattgctgacagcatggtctccctccctgattcagacataccagtiactgaa 1222
|||||
DB 306 GGGTGATTGCTGACAGCATGCTCTCCCTCCCTGATTCAGACATACACTTACTGAA 247
OY 1223 agcaaatcagtttaagtgattctcagtgctgaaaagcctgtccaggttccctccctt 1282
|||||
DB 246 AGCAATCAGTTTAAAGTATTCAGTGTGAGAAAGCCTGTCAGGTTCTCCCTT 187
OY 1283 tcccaagcctctctgttaataactcccttgggagaaactaacatcagtgccctcccgac 1342
|||||
DB 186 TCCCAAGCCTCTCTGTAAATCTCCCTTGGGCGAGCTACATCGGTGCTCCCGAC 127
OY 1343 ctgtgtaactagggacatgagagcaaaagagaggaagcaagccttgcttgagag 1402
|||||
DB 126 CTGTGACTAGGACATGGGAGCGCAAGAGGAGGAGCAAGCCCTGGCTGGCGAG 67
OY 1403 ttgtcaatggttggtggtgactgttctattttttaataaaaataaagatgagagaaa 1462
|||||
DB 66 TTGTCATGTGTTGGTGGTGAAGCTTTTATTTTAAATAAATAAAGATGAGAGAAA 7
OY 1463 ttaaaa 1468
|||||
DB 6 TTAATA 1

Search completed: February 28, 2001, 23:43:31
Job time: 19397 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 1, 2001, 02:41:45 ; Search time 8425.79 Seconds
(without alignments)
920.805 Million cell updates/sec

Title: US-09-602-597-3
1516
Sequence: 1 cgtccgagcctcgtgggga.....aaaaaaaaaaaaaaaa 1516

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 118133 segs, 2558875100 residues
Total number of hits satisfying chosen parameters: 2236266

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl:*
1: gb_bal:*
2: gb_bal:*
3: gb_bal:*
4: gb_bal:*
5: gb_bal:*
6: gb_bal:*
7: gb_bal:*
8: gb_bal:*
9: gb_bal:*
10: gb_bal:*
11: gb_bal:*
12: gb_bal:*
13: gb_bal:*
14: gb_bal:*
15: gb_bal:*
16: gb_bal:*
17: gb_bal:*
18: gb_bal:*
19: gb_bal:*
20: gb_bal:*
21: gb_bal:*
22: gb_bal:*
23: gb_bal:*
24: gb_bal:*
25: gb_bal:*
26: gb_bal:*
27: gb_bal:*
28: gb_bal:*
29: gb_bal:*
30: gb_bal:*
31: gb_bal:*
32: gb_bal:*
33: gb_bal:*
34: gb_bal:*
35: gb_bal:*
36: gb_bal:*
37: gb_bal:*
38: gb_bal:*
39: gb_bal:*
40: gb_bal:*
41: gb_bal:*
42: gb_bal:*
43: gb_bal:*

44: em_hcg2:*
45: em_hcg3:*
46: em_hcg4:*
47: em_hcg5:*
48: em_hcg6:*
49: em_hcg7:*
50: em_hcg8:*
51: em_hcg9:*
52: em_hcg10:*
53: em_hcg11:*
54: em_hcg12:*
55: em_hcg13:*
56: em_hcg14:*
57: em_hcg15:*
58: em_hcg16:*
59: em_hcg17:*
60: em_hcg18:*
61: em_hcg19:*
62: em_hcg20:*
63: em_hcg21:*
64: em_hcg22:*
65: em_hcg23:*
66: em_hcg24:*
67: em_hcg25:*
68: em_hcg26:*
69: em_hcg27:*
70: em_hcg28:*
71: em_hcg29:*
72: em_hcg30:*
73: em_hcg31:*
74: em_hcg32:*
75: em_hcg33:*
76: em_hcg34:*
77: em_hcg35:*
78: em_hcg36:*
79: em_hcg37:*
80: em_hcg38:*
81: em_hcg39:*
82: em_hcg40:*
83: em_hcg41:*
84: em_hcg42:*
85: em_hcg43:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1471	97.0	1505	10	AF137386 Homo sapi
2	1193.2	78.7	141132	77	HS192P9 Human DNA
3	934	61.6	221262	30	AC009090 Homo sapi
4	428.8	28.3	1475	11	RNPPLIPIN
5	428.8	28.3	1622	81	AS1771
6	416	27.4	1250	11	RNU13617
7	272	17.9	272	79	G06198
8	124.8	8.2	853	81	AR062277
9	78.4	5.2	168747	59	AC078781
10	78.4	5.2	179017	39	AC012520
11	78.4	5.2	195115	55	AC046140
12	74	4.9	1068	79	CNS01E0R
13	74	4.9	2337	78	HSMB00876
14	73.8	4.9	737	56	TCH2A
15	73.8	4.9	1288	56	TBBDPMSTN
16	73.6	4.9	3916	6	AF097728
17	73.6	4.9	67554	66	AC083962
18	73.2	4.8	1582	81	A77033
19	73.2	4.8	1582	81	A77033
20	73.2	4.8	1887	78	HSGLNACT
21	73.2	4.8	2896	4	AF061981


```

Db 931 GGGGATGTGAGGCTGTCTCTGCTGTCTTGTAGAGACTTCAAGCTCCAAAGCTGGG 990
Qy 963 cccaccccttccacacactaatgacatacaagaagactccagacttgagccccagac 1022
Db 991 CCCACCCCTTCTCACACACACTAAATGACATAACAGAGACTCCACACCTGAGCCCCGAC 1050
Qy 1023 ccgcgcagataagactcaacaagacacagctagacacttagcttggttccagagagc 1082
Db 1051 CCGCGCTAGATTAAGCTTAACACACACAGTACACCTTAGCTTTGTTCCAGAGAGC 1110
Qy 1083 tgcgaagctggtgaacacacactctccttcaaacacggttcaacacacactctcc 1142
Db 1111 TGAGCAGACTGTGTGAACACACTCTCTCTTTAAACACCGTTTCAACCAACCTCTCC 1170
Qy 1143 ggcgcacactgtataaagtgggttgatgtgtcagacagatgtctccctcctgac 1202
Db 1171 GGAAGCCAACTGTAAAAAGGGGTGATGCTACACACATGTCCTCTCCCTGCATTT 1230
Qy 1203 cagacatacactgactgaagaagaacacagctttaaagtgtctcagctgacgaagc 1262
Db 1231 CAGNCATACAGTTACTGAAAGCAATCACTTTTAAGTGTCTCTGAGTCTGAAAGCC 1290
Qy 1263 tgcgaagcttcccttcccttcccaagcctcctctgtataactcccttggcgaagc 1322
Db 1291 TGTCAGAGTTCTCTCTCTCCCAAGCCTCTCTGTATACCTCTTGGGGAAGCT 1350
Qy 1323 aacatgggtcctcccgacacttgctgactagacacatggagcgaagagagagaa 1382
Db 1351 AACATCGGTCTCTCCCGACCTTGTGACTGACATGAGCAGCAAGAGAGAGGAA 1410
Qy 1383 gcaagagctgctgctgagctgctgactgctggtgtgactggtttatattttaat 1442
Db 1411 GCAAGGCTTGGCTGGCGAATTGTCTATGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1470
Qy 1443 aaaaataaagatgagagaataaaaaaaa 1473
Db 1471 AAAAATAAAGATGAGAGAAATTAATAAAAAA 1501

RESULT 2
LOCUS HS192P9
DEFINITION Human DNA sequence from clone RP1-192P9 on chromosome Xp11.23-11.4
contains a pseudogene similar to rat Plasmolipin, ESTs and GSSs,
complete sequence.
ACCESSION AL020989
VERSION AL020989.2 GI:4539519
KEYWORDS HTG; Plasmolipin.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 141132)
AUTHORS Heath, P.
TITLE Direct Submission
JOURNAL Submitted (04-FEB-2000) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk; Clone
request: clonerequest@sanger.ac.uk
COMMENT On Mar 29, 1999 this sequence version replaced gi:4034132.
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unsure'
feature key.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
The following abbreviations are used to associate primary accession

```

numbers given in the feature table with their source databases:
 Em: EMBL; Sw: SWISSPROT; Tr: TrEMBL; Wp: WORMPEP; Information
 on the WORMPEP database can be found at
<http://www.sanger.ac.uk/Projects/C.elegans/wormpep> This sequence
 was generated from part of bacterial clone contigs of human
 chromosome X, constructed by the Sanger Centre Chromosome X
 Group. Further information can be found at
<http://www.sanger.ac.uk/HGP/ChrX>
 RP1-192P9 is from the library RPCT-1 constructed at the Roswell
 Park Cancer Institute by the group of Pieter de Jong. For further
 details see <http://bacpac.med.buffalo.edu/>
 VECTOR: pCYPAC2
 This sequence is the entire insert of clone RP1-192P9.

FEATURES

source

```

1..141132
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="X"
/map="p11.23-11.4"
/clone="RP1-192P9"
/clone_11b="RPCT-1"
1..3583
/feature="LIM1 repeat: matches 434..4056 of consensus"
3551..4648
/feature="LIP5 repeat: matches 4328..5453 of consensus"
5364..5741
/feature="LIM2 repeat: matches 5452..5829 of consensus"
6053..6531
/feature="LIM2 repeat: matches 5829..6308 of consensus"
9556..9628
/feature="MTL1A1-internal repeat: matches 1570..1651 of
consensus"
9627..9700
/feature="LIM1 repeat: matches 5442..5514 of consensus"
9703..11332
/feature="MTL1A1-internal repeat: matches 3..1561 of
consensus"
14098..14616
/feature="MER1A repeat: matches 1..527 of consensus"
14617..14888
/feature="LIM7 repeat: matches 5902..6169 of consensus"
15231..16025
/feature="LIM8 repeat: matches 5313..6159 of consensus"
16066..16404
/feature="LIM8 repeat: matches 4885..5213 of consensus"
15917..20048
/feature="LIM8 repeat: matches 271..402 of consensus"
20049..20575
/feature="MT2CA repeat: matches 1..501 of consensus"
20576..20907
/feature="LIM8 repeat: matches 402..734 of consensus"
20908..21045
/feature="LIM8 repeat: matches 6146..6283 of consensus"
22973..23705
/feature="LIP3 repeat: matches 5414..6147 of consensus"
25117..30634
/feature="LIP3 repeat: matches 1537..4181 of consensus"
30635..30885
/feature="LIP6 repeat: matches 5893..6143 of consensus"
30886..32124
/feature="LIP3 repeat: matches 4181..5403 of consensus"
34239..35400
/feature="LIP8 repeat: matches 4983..6154 of consensus"
41121..41996
/feature="THE1C-internal repeat: matches 1..928 of consensus"
44380..44610
/feature="Char1el repeat: matches 1..239 of consensus"
44598..44731
/feature="Char1el repeat: matches 2623..2761 of consensus"
48313..48597
/feature="LIP repeat: matches 4181..4465 of consensus"
49218..49531
/feature="MER45B repeat: matches 1..322 of consensus"

```



```
Db 115825 CTCCCTGGGGCTGAAGCCGCTACCTGGGTAGCAGACAGAGGTCACCCCGCAAGCCCAAA 115884
QY 726 gctgggagagccctcgctggagtcagcccaaggaactgcaatttcctctctgcgcgt 785
Db 115885 GCTTGGGAGAGCTTGAAGAGACTGAGCCCTGAGGAGACTACACTTGGTCTCTGCCCCAC 115944
QY 786 cagacataagctcacagcgtcgaagaagcagccagcgtgcagacatctcgcgttg 845
Db 115945 CAGACACAGCTCTCAGAGGCTTAAGAGACAGGCTGAGGATCTTGGGCTTG 116004
QY 846 caggaagcccaactgctgagacctctctccatcccccatttcagtgaaatgacggg 905
Db 116005 CAGGAGGCCAACCGCTGAGACCTCTCCCATCTCCCTTATTCAGTGAAGTTGAC-GGG 116063
QY 906 gactgagagctgtgctcctcctcctgtctttagaggaactcagcgtccaagaactggccc 965
Db 116064 GATCTAGGCTGTGCTCTCTCTCTCTCTCTTGAAGAGACTTCAAGGCTCCAGGCTGGGCCCC 116123
QY 966 accctctccacgaactaactgactcaacaagaactcagacactgaccccccagaccg 1025
Db 116124 ACCCTTCTCAGCAGACTAATATGATTAACAGAGACTTCAGACTGCAAGCCCAAGCCCA 116183
QY 1026 ccgtaagataagacctcaacaagcagcagctagcacttagctctgttcacagagagctga 1085
Db 116184 CTCTAGTATCAGGCTTAACAACAAGCATGAGCAGCTTCGCTTGTTCAGAGGAGCTGA 116243
QY 1086 gcaagctggtggaacacactctctctctttaaacaacgcttcaacacactctcctgta 1145
Db 116244 GCAAGCTCAAGAAACCACTTCTCTCTCTCTCTCTTAAGACACAAATTCGCCACCCCTCCCGGA 116303
QY 1146 gccacactgtaaaaagtgggt--tgattgtgcagcagatgctctcctccctgcgacttc 1203
Db 116304 GCCAAGCTGTAAAAAGCGGCTTATGATTGTTGACAGCATGCTTCATCCATCCGCAATTC 116363
QY 1204 agacataccagcttactgaaagcaaatcagtttaagtgatctcagtgctgaaagcct 1263
Db 116364 AGACATACAGCTTATTAAGAAAGCAATCCGTTTAAGTGAATGTCATGCTGAAGAACCT 116423
QY 1264 gtccaggttctctcctcctcccaagcctctctctgtaatactcctctggcggaagcta 1323
Db 116424 GTCCAGGTTTCTTCCCTTCTCCCAAGGCTCTCT--TGTAATATCTCCCTTTGGGCAAGCTTA 116482
QY 1324 acatc-----gtgctctccgacactgctactagcagacatggaagcgaagaggg 1376
Db 116483 ACATTCACCTGCCGAGCTGCTCCCTGACCTCTGACTAGGCAAGGAGGCAAGGAGGG 116542
QY 1377 aggaagcgaagccttgctcctgagcagctgcatgctgctgctgctgcttatt-- 1433
Db 116543 AGGGAAGCAAGGCTTCTTGGCCAGTTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 116602
QY 1434 -----tttttaataaataaagatgagaataaataaataaataaataaataa 1487
Db 116603 TGTGTTGTTTATTAATAAATAAGATGAGAGAAATTAATAAATAAAGACATCAAAATG 116662
QY 1488 aaaaaaataaataaataaataa 1507
Db 116663 CCAAGCCTATATGAAAAAA 116682
```

```
RESULT 3
AC009090 221262 bp DNA HTG 21-JUN-2000
DEFINITION Homo sapiens chromosome 16 clone RP11-407G23, WORKING DRAFT
ACCESSION AC009090
VERSION AC009090.7 GI:8575954
KEYWORDS HTG: HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 221262)
```

AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

DOE Joint Genome Institute.
Sequencing of Human Chromosome 16
Unpublished
2 (bases 1 to 221262)
DOE Joint Genome Institute.
Direct Submission
Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Jun 21, 2000 this sequence version replaced gi:7689913.
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: <http://www.jgi.doe.gov>

Project Information
Center Project Name: 562820
Center Clone Name: RP11-11_407G23

Summary Statistics
Consensus quality: 18861 bases at least Q40
Consensus quality: 204128 bases at least Q30
Consensus quality: 207087 bases at least Q20
Estimated insert size: 218870; agarose-fp estimation
Estimated insert size: 218762; sum-of-ctgts estimation
Quality coverage: 6.45 in Q20 bases; agarose-fp estimation
Quality coverage: 6.45 in Q20 bases; sum-of-ctgts estimation
NOTE: This is a 'working draft' sequence. It currently
consists of 26 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.
1
1225: contig of 1225 bp in length
1325: gap of unknown length
1326: contig of 1222 bp in length
2547: gap of unknown length
2647: gap of unknown length
4098: contig of 1451 bp in length
4198: gap of unknown length
4199: contig of 1996 bp in length
6194: gap of unknown length
6295: 7398: contig of 1104 bp in length
7399: gap of unknown length
7498: gap of unknown length
7499: 8890: contig of 1392 bp in length
8890: gap of unknown length
8891: 10074: contig of 1084 bp in length
10074: gap of unknown length
10175: 11422: contig of 1248 bp in length
11422: gap of unknown length
11423: 11522: gap of unknown length
11523: 13032: contig of 1510 bp in length
13033: 13132: gap of unknown length
13133: 14242: contig of 1110 bp in length
14242: gap of unknown length
14343: 15547: contig of 1205 bp in length
15548: 15647: gap of unknown length
15648: 17286: contig of 1635 bp in length
17287: 17387: gap of unknown length
17387: 18969: contig of 1583 bp in length
18970: 19069: gap of unknown length
19070: 21434: contig of 2365 bp in length
21435: 21534: gap of unknown length
21535: 23455: contig of 1921 bp in length
23456: 23555: gap of unknown length
24783: 24783: contig of 1228 bp in length
24784: 24883: gap of unknown length
24884: 27886: contig of 3003 bp in length
27887: 27986: gap of unknown length
27987: 30722: contig of 2736 bp in length
30723: 30822: gap of unknown length
30823: 35202: contig of 4380 bp in length
35203: 35302: gap of unknown length
35303: 45765: contig of 10463 bp in length


```

/dev_stage="adult"
/tissue_type="kidney"
64. .537
CDS

```

BASE COUNT	243	360	332	315
ORIGIN	a	c	g	t

Query Match	27.4%;	Score 416;	DB 11;	Length 1250;
Best Local Similarity	83.3%;	Pred. No. 1.7e-61;		
Matches 473;	Conservative 0;	Mismatches 95;	Indels 0;	Gaps 0;

QY 113 ccgtcgaagttagcacgcgcgacccagtcagctccctcgccagggcccgaaacctctgctgctc 172
Db 1 CCCTCGAAAGTGACGACCAGGACCAGCAGCCGCCGGCAGGGCGTGGAGCATCTGTGTCT 60

QY 173 ggcgtcgcgcccgacctggcttcgtcgtcccccctcctgggagcgtacatgctctgcag 232
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 61 GCGATGCGACCGAGACTGGGGCTTGTGCGCTCCGCCCTTGGGGTGCTCGCGCTTCGACG 120

233 ctgtgtcctgggagctctgtgtgtgtggcgctgattgcgcgaacccgtaccacactgtatccg 292
 |||||
 121 ctggtgctgggggctgctggtgtgtgggacctgtgattgctgacaccccatcaccactgtatcct 180

293 gccatcgcgcggcgtatgttcgcgcgcgtctctctcgtgcgcgcacatgcctcttc 352
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 181 GCCACGCGCGGGTATGTTTGTAGCTCTTCTCTGGCTGGTACAAATTGCTCTTC 240

Dy 353 aaactctacctgtttcagctgcacatgaagtgttaccatggttcctcgccaactgytgta 412
+ + + + +
Db 241 ATCATCTACCTGTTTCACATGCACATGAAGTGTATATGGTGCCCTGGCCGTTGTGTTTA 300

```

413 atgactcttaacatcacgagccacgcgtctctacatcacgcctcatgcctgcgcg 472
      || ||||| | ||||| ||||| ||||| ||||| | |||
Db 301 CTGCTCTTCTTTGTTGCTGCCACCGTCTCTATATACACCGCTTTGTGCGCCTGTGCGGCG 360

```

[illegible]

```

QY 533 ttctctgcgttttgatgatgcctatgagtgagtcctcttcagctacacgagcc 592
      ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 421 ttctttgcttgcctgcatgatgtgcctacggcactgagtgctttcttcacgcttccacgccc 480

```

Qy 593 tggcgaagagtagcgaacatgcgcaccacagtcagatgctgcgcgtatgcctaacc 652
 |||||
 Db 481 TGGCGAGGAGTGGCGACGAACGCGGCGCACGAGTCAGATGGCTGGGGGCTTCTTAAGCC 540
 |||||

Qy	653	acctgtgcacagccccctctgggagctg	680
Db	541	AGCTATGCTGTGACCTAAGCCACGGCTG	568

RESULT	7
G06198/c	
LOCUS	
G06198	
272 bp	
DNA	
SMS	
19-OCT-1995	

DEFINITION	human STS WI-62/5.
ACCESSION	G06198
VERSION	G06198.1 GI:859443
KEYWORDS	
STS sequence: primer:	sequence tagged site

SOURCE	ORGANISM
human SVSs derived from sequences in dbEST and the Unigene collection.	Homo sapiens
	Eukaryotae: mitochondrial eukaryotes. Metazoa: Chordata:

Catarrhini; Homnidae; Homo.
Tetrapoda; Gnathostomata; Osteichthyes; Sarcopterygii; Chonata;
Vertebrata; Amniota; Mammalia; Theria; Eutheria; Archonta; Primates;
Catarrhini; Homnidae; Homo.

REFERENCE	1 (bases 1 to 272)
AUTHORS	Hudson,T.
TITLE	Whitehead Institute/MIT Center for Genome Research: Physically Mapped ESTs
JOURNAL	Unpublished (1995)
COMMENT	

Contact: Thomas Hudson
Whitehead Institute/MIT Center for Genome Research
Whitehead Institute for Biomedical Research
9 Cambridge Center, Cambridge MA 02142 USA
Tel: 617 252 1900
Fax: 617 252 1902
Email: thudson@genome.wi.mit.edu

Primer A: AACAGTCACCACCACACCACA
Primer B: TTCAGACATACCAGTTACTGAAGC
Sns size: 228
PCR Profile:

Denaturation:
Annealing: 56 degrees C
Polymerization:
PCR Cycles: 35
Thermal Cycler:
Protocol:
Incubator: 10.0

Template: 10 ng
Primer: each 5 μ M
dNTPs: each 4 nM
Taq Polymerase: 0.025 units/ μ l
Total Vol: 20 μ l

Bulter:
MgCl₂: 1.5 mM
KCl: 50 mM
Tris-HCl: 10 mM
pH: 9.3

Prepared with primer pairs derived from Z38289 -- dbEST.
Location/Qualifiers

	/organism="Homo sapiens"	
STS	37. .264	
primer_bind	37. .56	
primer_bind	complement(240. .264)	
BASE COUNT	79 a 64 c 61 g 68 t	
ORIGIN		

Query Match	17.9%	Score 272:	DB 79:	Length 272:
Best Local Similarity	100.0%	Pred. No.	6.2e-37:	
Matches 272:	Conservative 0:	Mismatches 0:	Indels 0:	Gaps 0:

[illegible][illegible]

Db 152 GGGCGAAGCTAACATCGGTGCTCCCTCCGACCTTGCTGCATGAGCACATGGCAGCAAAAGG 93

Dy 1373 agggaggaagaacgagccctgtccgcgcatlgtcattgttgttgatgcgttttat 1382
|||||
Db 92 AGGAGGAAGCAAGCCCTTGCCTGGCGAATTGATGTGGTTGGTCGTACTTTPAT 33

```
Qy      1433 ttttlttaaaaaataagaatgagagaatt 1464  
        |||||  
Db      32 TTTTAAATAAAAAATTAAGATGAGAGAAATT 1
```


[illegible]

ORIGIN

Query Match 4.9%; Score 73.8; DB 56; Length 1288;
 Best Local Similarity 87.1%; Pred. No. 0.0037;
 Matches 81; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

OY 1424 ctgttttattttttaataaagatgaggaatttaaaaaaaaaaaaaaaaa 1483
 ||| ||||||||| ||| ||| ||| ||||||||| |||
 Db 1171 CCGTCTATTTTAAATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1230

OY 1484 aaaaaaaaaaaaaaaaaaaaaaaaaa 1516
 ||||||||| ||||||||| ||||||||| |||
 Db 1231 AAAAAAAAAAAAAAAAAAAAAAAAAA 1263

Search completed: March 1, 2001, 02:52:53
 Job time: 25779 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 1, 2001, 02:11:11 ; Search time 234.97 Seconds
(without alignments)
2423.733 Million cell updates/sec

Title: US-09-602-597-3

Sequence: 1 cgtccgagcgtcgtgggggaa.....aaaaaaaaaaaaaaaaaa 1516

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 480022 seqs, 187831343 residues

Total number of hits satisfying chosen parameters: 960044

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

N_Geneseq_36: *
1: /SID56/gcgdata/geneseq/geneseqn/NA1980.DAT: *
2: /SID56/gcgdata/geneseq/geneseqn/NA1981.DAT: *
3: /SID56/gcgdata/geneseq/geneseqn/NA1982.DAT: *
4: /SID56/gcgdata/geneseq/geneseqn/NA1983.DAT: *
5: /SID56/gcgdata/geneseq/geneseqn/NA1984.DAT: *
6: /SID56/gcgdata/geneseq/geneseqn/NA1985.DAT: *
7: /SID56/gcgdata/geneseq/geneseqn/NA1986.DAT: *
8: /SID56/gcgdata/geneseq/geneseqn/NA1987.DAT: *
9: /SID56/gcgdata/geneseq/geneseqn/NA1988.DAT: *
10: /SID56/gcgdata/geneseq/geneseqn/NA1989.DAT: *
11: /SID56/gcgdata/geneseq/geneseqn/NA1990.DAT: *
12: /SID56/gcgdata/geneseq/geneseqn/NA1991.DAT: *
13: /SID56/gcgdata/geneseq/geneseqn/NA1992.DAT: *
14: /SID56/gcgdata/geneseq/geneseqn/NA1993.DAT: *
15: /SID56/gcgdata/geneseq/geneseqn/NA1994.DAT: *
16: /SID56/gcgdata/geneseq/geneseqn/NA1995.DAT: *
17: /SID56/gcgdata/geneseq/geneseqn/NA1996.DAT: *
18: /SID56/gcgdata/geneseq/geneseqn/NA1997.DAT: *
19: /SID56/gcgdata/geneseq/geneseqn/NA1998.DAT: *
20: /SID56/gcgdata/geneseq/geneseqn/NA1999.DAT: *
21: /SID56/gcgdata/geneseq/geneseqn/NA2000.DAT: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	842.8	55.6	884	21	Z51624	Human membrane cha
2	428.8	28.3	1622	17	T34592	NTII-11 nerve prot
3	294.6	19.4	312	16	T22456	Human gene signatu
4	227.6	15.0	228	19	X10969	Human biallelic po
5	125	8.2	1152	21	Z36249	cDNA encoding a bo
6	124.8	8.2	853	19	V17143	Human proteolipid
7	108.4	7.2	600	20	Z24894	Human secreted pro
8	73.2	4.8	1582	17	T28259	Survival motor neu
9	73.2	4.8	1582	17	T18831	Human survival mot
10	72.2	4.8	1114	19	V34315	Human secreted pro
11	72.2	4.8	10288	18	T71322	Plasmid PCB51 enco
12	72.2	4.8	13414	18	T71321	Plasmid PCB50 enco

13	71.6	4.7	773	20	X37388	Human secreted pro
14	71.6	4.7	1474	18	T90174	Oil seed rape cyst
15	71.4	4.7	641	21	Z33356	Human secreted pro
16	71.2	4.7	99	18	T91300	Human M97-2 secret
17	71.2	4.7	101	18	V00420	3' fragment of clo
18	71.2	4.7	1037	21	A26447	Human secreted pro
19	71	4.7	281	21	A00286	Human colon cancer
20	70.8	4.7	1052	21	A26358	Human secreted pro
21	70.8	4.7	1813	16	Q88760	Human ubiquitous n
22	70.8	4.7	1898	18	T79634	DNA encoding human
23	70.6	4.7	144	19	V21239	Homo sapiens clone
24	70.6	4.7	144	20	X77331	Human secreted pro
25	70.4	4.6	485	21	Z65037	Membrane bound pro
26	70.4	4.6	1361	21	A26399	Human secreted pro
27	70	4.6	1034	21	Z52527	Human secreted pro
28	69.6	4.6	1933	21	A23442	Human secreted pro
29	69.4	4.6	1376	20	X04325	Human secreted pro
30	69.4	4.6	2475	21	A37059	Human secreted pro
31	69.2	4.6	2377	13	O20360	Human PRO1246 (UNG
32	69.2	4.6	2427	11	O04107	Human pro-urokinas
33	69	4.6	3410	19	V61201	Human pro-urokinas
34	69	4.6	3410	19	V58586	Prostate tumour sp
35	69	4.6	3410	21	A06349	Human immunogenic
36	69	4.6	3535	21	Z51555	Human hypoxia resp
37	69	4.6	3820	11	O03517	Plasmid pfcH1 enco
38	69	4.6	11326	19	V12373	Chimeric protease
39	68.8	4.5	886	20	X51754	DNA encoding a hum
40	68.8	4.5	1299	21	A16666	Human secreted pro
41	68.8	4.5	1812	21	A16649	Human secreted pro
42	68.6	4.5	84	19	V09269	Nucleotide sequenc
43	68.6	4.5	84	19	V04275	Secreted protein C
44	68.6	4.5	997	20	X52274	Protein PRO244 con
45	68.6	4.5	1118	13	Q20263	Deg-1 gene. Caeno

ALIGNMENTS

RESULT	1	
ID	Z51624	standard; cDNA; 884 BP.
XX		
DT	21-JUN-2000	(first entry)
XX		
DE	Human membrane channel protein-8 (MECHP-8) cDNA.	
XX		
KW	Membrane channel protein-8; MECHP-8; diagnosis; treatment; lymphoma;	
KW	cell proliferative disorder; bursitis; atherosclerosis; cancer; sarcoma;	
KW	inflammatory disorder; AIDS; Addison's disease; cystic fibrosis; asthma;	
KW	diabetes mellitus; osmoregulatory disorder; diarrhoea; renal failure;	
KW	muscular disorder; myocardiitis; Duchenne's muscular dystrophy; nocturnal;	
KW	cardiovascular disorder; hypertension; bronchitis; vasculitis; cardiac;	
KW	neurological disorder; Alzheimer's disease; Parkinson's disease; human;	
KW	Huntington's disease; antiarteriosclerotic; hepatotropic; cytostatic;	
KW	anti-HIV; antianaemic; neuroprotective; immunomodulator; antidiabetic;	
KW	hypotensive; vasotropic; antinflammatory; antiinflammatory; antidepressant;	
KW	anticonvulsant; thrombolytic; antiParkinsonian; immunostimulant; ss.	
XX		
OS	Homo sapiens.	
XX		
EH	Key	Location/Qualifiers
FT	CDS	120..668
FT		/tag= a
FT		/product= "MECHP-8"
FT	misc_binding	153..197
FT		/tag= b
FT		/bound_moiety= "Primer or Probe"
XX		
PN	WO2000012711-A2.	
XX		
PD	09-MAR-2000.	

XX 02-SEP-1999; 99WO-US20468.
 PF 02-SEP-1998; 98US-0145815.
 XX 12-NOV-1998; 98US-0191283.
 PR 09-DEC-1998; 98US-0208821.
 PR 26-JAN-1999; 99US-0237506.
 PR 10-FEB-1999; 99US-0247891.
 XX (INCYTE) INCYTE PHARM INC.
 PA Au-Young J, Bandman O, Tang Y, Reddy R, Hillman JL, Yue H;
 PI Lal P, Corley NC, Guegler KJ, Gorgone G, Baughn MR, Azimzal Y;
 XX WPI: 2000-256643/22.
 DR P-PSDB: Y70458.
 DR Novel human membrane channel protein and polynucleotide useful for
 PT diagnosing and treating cell proliferative, inflammatory, secretory,
 PT osmoregulatory, muscular, cardiovascular and neurological disorders
 XX
 PS Claim 9; Page 123; 140pp; English.
 CC The present sequence is a cDNA identified in Incyte clone 1375415
 CC derived from LUNGNOT10 cDNA library. It encodes human membrane channel
 CC protein-8 (MECHP-8), which is expressed in nervous and gastrointestinal
 CC tissues. Anti-MECHP antibodies can be used as therapeutic antagonists and
 CC reagents for diagnosis and monitoring diseases. MECHP cDNA can be used
 CC for diagnosis of MECHP-related diseases and gene mapping. MECHP can be
 CC used for treatment of cell proliferative disorders such as bursitis and
 CC atherosclerosis, cancers like lymphoma and sarcoma, inflammatory
 CC disorders like AIDS and Addison's disease, transport/secretory disorders
 CC like cystic fibrosis and diabetes mellitus, osmoregulatory disorders like
 CC diarrhoea and renal failure, muscular disorders like myocarditis and
 CC Duchenne's muscular dystrophy, cardiovascular disorders like hypertension
 CC and vasculitis, congenital lung anomalies like bronchitis and asthma and
 CC neurological disorders like Alzheimer's disease, Parkinson's disease and
 CC Huntington's disease.
 XX
 XX Sequence 884 BP; 146 A; 280 C; 282 G; 173 T; 3 other;
 SO
 Query Match 55.6%; Score 842.4; DB 21; Length 884;
 Best Local Similarity 99.5%; Pred. No. 1e-133;
 Matches 854; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

QY 423 acatcagccacccgtctctacatcacccgctccatccgctctctcgcggaattgacc 482
 DB 442 acatcagccacccgtctctacatcacccgctccatccgctctctcgcggaattgacc 501
 QY 483 tgacatccctgaagggaacccgctctataacacgagcgagctgctctctctctc 542
 DB 502 tgacatccctgaagggaacccgctctataacacgagcgagctgctctctctctc 561
 QY 543 gtttgatgatgatgacctgagtgagtgctctctctacacgagcctggagagag 602
 DB 562 gtttgatgatgatgacctgagtgagtgctctctctacacgagcctggagagag 621
 QY 603 tagcagcaatgcgagccacccagtcagatgctgagcgtacatgctaaacacccgtgcca 662
 DB 622 tagcagcaatgcgagccacccagtcagatgctgagcgtacatgctaaacacccgtgcca 681
 QY 663 cggcccccctctggggtgaaagcgcgcgtggtgacagagagagagagagagagag 722
 DB 682 cggcccccctctggggtgaaagcgcgcgtggtgacagagagagagagagagagag 741
 QY 723 gaagctgggagagccctggtgagtgacagccacagagagagagagagagagagag 781
 DB 742 gaagctgggagagccctggtgagtgacagccacagagagagagagagagagagag 801
 QY 782 ccgtcagacataagctctcagacgctaaagagagagagagagagagagagagagagag 841
 DB 802 ccgtcagacataagctctcagacgctaaagagagagagagagagagagagagagagag 861
 QY 842 ctgcag 859
 DB 862 ctgcag 879
 RESULT 2
 T34592
 ID T34592 standard; cDNA; 1622 BP.
 XX T34592; .
 AC
 XX 01-JAN-1997 (first entry)
 DE NTII-11 nerve protein coding sequence.
 DE
 XX Probe; central nervous system; peripheral nervous system; CNS; PNS;
 KW axon; neurone; neuronal cells; glial cells; Schwann cells; trauma;
 KW pathology; regeneration; MS2; metalloproteinase; ds.
 KW
 OS Homo sapiens.
 XX
 XX
 FH Key Location/Qualifiers
 FT mat_peptide 362..907
 FT /tag= a
 FT /product= NTII-11 protein expressed in damaged or
 FT regenerating nerve cells.
 XX
 XX MO9617865-A2.
 XX
 XX PD 13-JUN-1996.
 XX
 XX PF 05-DEC-1995; 95WO-EP04777.
 XX
 XX PR 27-JAN-1995; 95DE-4002525.
 XX PR 05-DEC-1994; 94DE-4443159.
 XX
 XX (BOE) BOEHRINGER MANNHEIM GMBH.
 XX
 XX PI Gillen C, Gleichmann M, Mueller H;
 XX
 XX WPI: 1996-287114/29.
 XX DR P-PSDB: R99799.
 XX
 PT Nucleic acids associated with damaged or regenerating nerve cells

|||||
Db 181 gcaagctgtgaacacactctccctccttaaacancgttcaacanccttncntgga 240
Qy 116 gccaccgttaaaagtggtgattgctgcagacagatggtctccctccctcattgag 1205
Db 241 gccacacgttaaaaagtggtgattgctgcagacagatggtctccctccctcattgag 300
Qy 1206 acataccagtt 1216
Db 301 acataccagtt 311
RESULT 4
X10969/c
ID X10969 standard; DNA; 228 BP.
XX X10969;
AC X10969;
XX 30-MAR-1999 (first entry)
DE Human biallelic polymorphic DNA fragment WI-6275.
XX
KM Polymorphism; biallelic; human; forensic; paternity testing; disease;
KM detection; phenotypic typing; characteristic; infection; hereditary;
KM autoimmune disease; cancer; inflammation; drug; therapy; medication;
KM treatment; marker; ss.
XX
OS Homo sapiens.
XX
FN W09820165-A2.
XX 14-MAY-1998.
PD
XX 05-NOV-1997; 97WO-US20313.
PF
XX 06-NOV-1996; 96US-0030455.
PR
XX (WRED) WHITEHEAD INST BIOMEDICAL RES.
XX
PI Hudson T, Lander ES, Wang D;
XX
DR WPI: 1998-286974/25.
XX
PT New isolated nucleic acid segments from the human genome - used for
PT determining polymorphic forms for use in e.g. forensics, paternity
PT testing or phenotypic typing for disease
XX
PS Claim 1; Page 101; 310pp; English.
XX
CC X10269-X12937 are human DNA fragments which contain biallelic polymorphic
CC markers which have been isolated using the primers represented in
CC X09121-X10268. The base occupying the polymorphic site is indicated by
CC the appropriate IUPAC-IUB ambiguity code. These fragments can be used in
CC methods for determining polymorphic forms in an individual for use in
CC e.g. forensics, paternity testing or for phenotypic typing for diseases
CC such as agammaglobulinemia, diabetes insipidus, Lesch-Nyhan syndrome,
CC muscular dystrophy, Wiskott-Aldrich syndrome, Fabry's disease, familial
CC hypercholesterolemia, polycystic kidney disease, hereditary
CC spherocytosis, von Willebrand's disease, tuberous sclerosis, hereditary
CC haemorrhagic telangiectasia, familial colonic polyposis, Ehlers-Danlos
CC syndrome, osteogenesis imperfecta, acute intermittent porphyria,
CC autoimmune diseases, inflammation, cancer, diseases of the nervous
CC system, infection by pathogenic microorganisms, and characteristics such
CC as longevity, appearance (e.g. baldness, obesity), strength, speed,
CC endurance, fertility, and susceptibility or receptivity to particular
CC drugs or therapeutic treatments. The isolated polymorphic nucleic acid
CC segments can also be used to produce medicaments for the treatment or
CC prophylaxis of such diseases.
XX
SQ Sequence 228 BP; 62 A; 59 C; 56 G; 50 T; 1 other;

Query Match 15.0%; Score 227.6; DB 19; Length 228;

Best Local Similarity 99.6%; Pred. No. 2.4e-30;
Matches 227; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
Qy 1201 ttcaacatcaccaattactgaagaacaaatcagtttaagtgtatctcgaagctgaag 1260
Db 228 ttccagacatcaccaattactgaagaacaaatcagtttaagtgtatctcgaagctgaag 169
Qy 1261 cctgtccaggttctcctccctcccaagccctctctctgtaatactcccttggcgaag 1320
Db 168 cctgtccaggttctcctccctcccaagccctctctctgtaatactcccttggcgaag 109
Qy 1321 ctacatcggtgctcccccagacctgtgctgactaggacacatggagcgaagaggag 1380
Db 108 ctacatcggtgctcccccagacctgtgctgactaggacacatggagcgaagaggag 49
Qy 1381 aagcaagccttgctgctgaggtgtcatgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1428
Db 48 aagcaagccttgctgctgaggtgtcatgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1
RESULT 5
Z36249
ID Z36249 standard; cDNA; 1152 BP.
XX Z36249;
AC Z36249;
XX 22-FEB-2000 (first entry)
DE cDNA encoding a bone marrow secreted protein designated BMS61.
XX
XX Bone marrow secreted protein; bone marrow stromal cell; cytokine;
KM cell proliferation; cell differentiation; hematopoiesis; anaemia;
KM myeloid cell deficiency; lymphoid cell deficiency; myeloid cell;
KM erythroid progenitor cell; colony stimulating factor; granulocyte;
KM monocyte; macrophage; myelo-suppression; megakaryocyte; platelet;
KM platelet disorder; thrombocytopenia; hematopoietic stem cell;
KM stem cell disorder; aplastic anaemia; bone differentiation;
KM paroxysmal nocturnal hemoglobinuria; bone growth; cartilage; tendon;
KM ligament; nerve; wound healing; tissue repair; burn; incision; ulcer;
KM bone fracture; cartilage damage; artificial joint; ss.
XX
XX Homo sapiens.
OS
XX
XX Key Location/Qualifiers
XX FH 276..797
XX FT CDS /*tag=a
XX FT /product= "bone marrow secreted protein"
XX
XX W09933979-A2.
XX
XX 08-JUL-1999.
XX
XX 18-DEC-1998; 98WO-US27008.
PF
XX 30-DEC-1997; 97US-0068958.
PR 24-SEP-1998; 98US-0101603.
PR 30-SEP-1998; 98US-0102540.
XX
XX (CHIR) CHIRON CORP.
XX
XX Lin H, Cao L;
XX
XX WPI: 2000-038344/03.
DR P-PSDB: Y53643.
XX
XX New isolated human polynucleotide and secreted proteins can induce
PT production of other cytokines in certain cell populations -
XX
XX Claim 11; Page 118-119; 120pp; English.
XX
CC Z36228-49 encode bone marrow secreted proteins of human bone marrow
CC stromal cells. The proteins can exhibit cytokine, cell proliferation, or
CC cell differentiation activity (either inducing or inhibiting). They can


```

XX 19-OCT-1994; 94EP-0402353.
XX
XX (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.
XX
XX Melk1 J, Munnich A;
XX
XX WPI; 1996-202055/21.
XX
XX New isolated survival motor neuron gene - used to develop prods. for
XX the diagnosis and treatment of motor neuron diseases
XX
XX Claim 4; Fig 2a; 29pp; English.
XX
XX The SMN gene is a chromosome 5-SMA (Spinal Muscular Atrophy)
XX determining gene. Sequences of clone T-BCD541 and clone C-BCD541
XX are provided in T28255 and T28259, respectively.
XX SMN sequences can be used for detecting neuron disorders having
XX symptoms of muscular weakness with or without sensory changes
XX such as amyotrophic lateral sclerosis (ALS), SMA, primary lateral
XX sclerosis (PLS), etc.
XX
XX Sequence 1582 BP; 562 A; 285 C; 335 G; 400 T; 0 other;
XX
Query Match
Best Local Similarity 86.2%; Score 73.2; DB 17; Length 1582;
Matches 81; Conservative 0; Mismatches 13; Indels 0; Gaps 0;
XX
QY 1423 acgtttatttttttaataaataagatgagaaatttaaaaaaaaaaaaaa 1482
DB 1473 aatatttaatttttttaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 1532
QY 1483 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 1516
DB 1533 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 1566
XX
RESULT 9
TID 118831
XX 118831 standard; cDNA; 1582 BP.
XX
AC 118831;
XX
DT 02-OCT-1996 (first entry)
XX
DE Human survival motor neuron variant gene cDNA clone C-BCD541.
XX
XX Survival motor neuron gene; SMN gene; spinal muscular atrophy;
XX chromosome 5-SMA determining gene; amyotrophic lateral sclerosis;
XX primary lateral sclerosis; arthrogryposis multiplex congenita;
XX diagnosis; gene therapy; C-BCD541; ss.
XX
OS Homo sapiens.
XX
XX Key Location/Qualifiers
FH CDS 34..918
FT /tag- a
FT 1..114
FT /tag- b
FT /label= Exon-1
FT 115..306
FT /tag- c
FT /label= Exon-2
FT 307..507
FT /tag- d
FT /label= Exon-3
FT 508..660
FT /tag- e
FT /label= Exon-4
FT 661..756
FT /tag- f
FT /label= Exon-5
FT

```

```

FT exon 757..867
FT /tag- g
FT /label= Exon-6
FT exon 868..921
FT /tag- h
FT /label= Exon-7
FT variation 873
FT /tag- i
FT /note= "base 873 (c) is c is T-BCD541"
FT exon 922..1483
FT /tag- j
FT /label= Exon-8
FT variation 1155
FT /tag- k
FT /note= "base 1155 (a) is g in T-BCD541"
XX
XX EF711833-A2.
XX
XX 15-MAY-1996.
XX
XX 19-OCT-1995; 95EP-0402335.
XX
XX 19-OCT-1994; 94EP-0402353.
XX
XX (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.
XX
XX Melk1 J, Munnich A;
XX
XX WPI; 1996-232098/24.
XX
XX Human survival motor neuron gene T-BCD541, variant C-BCD541 and
XX murine equv. - useful to develop primers and probes for in vitro
XX detection of motor neuron diseases e.g. spinal muscular atrophy
XX
XX Claim 4; Fig 2; 47pp; English.
XX
XX A cDNA sequence (T18831) was detd. for the human survival motor
XX neuron (SMN) variant gene C-BCD541. This centromeric gene
XX differs from the SMN telomeric gene cDNA clone T-BCD541 (T18828)
XX by 2 nucleotides. The C-BCD541 gene is unable to correct in
XX vivo a deficiency of the T-BCD541 gene responsible for motor
XX neuron diseases of the spinal muscular atrophy type. A variation
XX of sequence within intron 7 of the C-BCD541 genomic sequence
XX results in an alternatively spliced transcript that lacks exon 7
XX and a truncated SMN protein (R94965) lacking the exon 7-encoded
XX C-terminal region.
XX
XX Sequence 1582 BP; 562 A; 283 C; 335 G; 402 T; 0 other;
XX
Query Match
Best Local Similarity 86.2%; Score 73.2; DB 17; Length 1582;
Matches 81; Conservative 0; Mismatches 13; Indels 0; Gaps 0;
XX
QY 1423 acgtttatttttttaataaataagatgagaaatttaaaaaaaaaaaaaa 1482
DB 1473 aatatttaatttttttaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 1532
QY 1483 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 1516
DB 1533 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 1566
XX
RESULT 10
V34315
ID V34315 standard; DNA; 1114 BP.
XX
AC V34315;
XX
DT 29-JAN-1999 (first entry)
XX
DE Human secreted protein gene 5 clone HELD941.
XX

```

KW Human: secreted protein; fusion protein; gene therapy; protein therapy;
diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;
developmental abnormality; foetal deficiency; blood; allergy; renal; ds;
immune system; asthma; lymphocytic disease; brain; hepatitis; lymphoma;
inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;
cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.
XX
OS Homo sapiens.
XX
FN W09840483-A2.
XX
PD 17-SEP-1998.
XX
PE 12-MAR-1998; 98WO-US04858.
XX
PR 19-DEC-1997; 97US-0068368.
PR 14-MAR-1997; 97US-0040710.
PR 14-MAR-1997; 97US-0040762.
PR 30-MAY-1997; 97US-0048100.
PR 30-MAY-1997; 97US-0048189.
PR 30-MAY-1997; 97US-0048357.
PR 30-MAY-1997; 97US-0050934.
PR 06-JUN-1997; 97US-0048970.
PR 05-SEP-1997; 97US-0057765.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Ferrie AM, Fischer CL, Gentz RL, Greene JM, Kyaw H;
PI Li H, Li Y, Moore PA, Rosen CA, Ruben SM, Soppet DR;
PI Wei YF, Young PE, Zeng Z;
XX
DR WPI: 1998-520811/44.
DR P-PSDB; W75225.
XX
PT Isolated human poly:nucleotide(s) encoding secretory peptide(s) -
PT used to develop products for the diagnosis and treatment of e.g.
PT inflammation, cancers, CNS disorders or immune system disorders
XX
PS Claim 1; Page 142-143; 201pp; English.
XX
CC This sequence represents a nucleic acid molecule which encodes a secreted
CC human protein. The gene number, and the clone it is derived from, are
CC detailed in the descriptor line. The gene can be used to generate fusion
CC proteins by linking to the gene to a human immunoglobulin Fc portion
CC (e.g. V34277) for increasing the stability of the fused protein as
CC compared to the human protein only.
CC The invention relates to 28 novel genes and their fragments (nucleic acid
CC sequences: V34286-V34325; amino acid sequences W75196-W75235) which
CC are useful for preventing, treating or ameliorating medical conditions
CC e.g. by protein or gene therapy. Also, pathological conditions can be
CC diagnosed by determining the amount of the new polypeptides in a sample
CC or by determining the presence of mutations in the new polynucleotides.
CC Specific uses are described for each of the 28 polynucleotides, based on
CC which tissues they are most highly expressed in (see V34286 for described
CC uses).
XX
SQ Sequence 1114 BP; 371 A; 196 C; 272 G; 272 T; 3 other;

Query Match 4.8%; Score 72.2; DB 19; Length 1114;
Best Local Similarity 80.8%; Pred. No. 0.00041;
Matches 80; Conservative 3; Mismatches 16; Indels 0; Gaps 0;

OY 1418 tggtagcgtttatttttttaataaataagatgaggaatttaaaaaaaaaa 1477
DB 1015 kgttggctkttttatttttaaaaaaaaaaaaaaaaaaaaaaaaaa 1074
OY 1478 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 1516
DB 1075 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 1113

RESULT 11
ID T71322/c
XX T71322 standard; DNA: 10288 BP.
XX
AC T71322;
DE 11-SEP-1997 (first entry)
XX
DE Plasmid pCB51 encoding partial UNC-53 protein used as bait vector.
XX
KW UNC-53; neuronal regeneration; revascularisation; wound healing;
KW neurodegenerative disease; Alzheimer's disease; Huntington's;
KW peripheral neuropathies; metastasis inhibition; cancer;
KW circular; Caenorhabditis elegans; ds.
XX
OS Synthetic.
XX
FN W09638555-A2.
XX
PD 05-DEC-1996.
XX
PE 31-MAY-1996; 96WO-EP02311.
XX
PR 31-MAY-1995; 95GB-0010944.
XX
PA (BOGA/) BOGAERT T.
PA (STRID/) STRINGHAM E.
PA (VAND/) VANDEKERCKHOVE J.
XX
PI Bogaert T, Stringham E, Vandekerckhove J;
XX
DR WPI: 1997-034369/03.
XX
PT Caenorhabditis elegans UNC-53 protein 8A and 7A variants - useful to
PT promote neuronal regeneration, revascularisation or wound healing.
XX
PS Example 15; Page 139-145; 278pp; English.
XX
CC T71321 and T71322 are plasmids pCB50 and pCB51, respectively. They
CC are used as bait vectors in a yeast two hybrid system. pCB50 encodes
CC the full length UNC-53 protein variant 7A of Caenorhabditis elegans
CC and pCB51 encodes a C-terminal portion of this protein. Both vectors
CC were constructed by cloning the appropriate cDNA fragment into
CC pAS1-CH2 (Clontech). The yeast two hybrid system was set up for
CC identifying proteins that interact with UNC-53. UNC-53 protein 8A and
CC 7A variants and nucleic acids encoding them are useful as medicaments
CC to promote neuronal regeneration, revascularisation or wound healing, or
CC for treatment of chronic neurodegenerative diseases (e.g. Alzheimer's or
CC Huntington's disease) or acute traumatic injuries. Transgenic cells and
CC organisms transfected with UNC-53 cDNA can be used to determine whether
CC a substance is an inhibitor or enhancer of the regulation of cell shape
CC or motility or the direction of cell migration by screening for a
CC phenotypic change in the cell. Inhibitors can be used to alleviate the
CC spread of disease inducing cells or metastasis. Probes derived from the
CC cDNA sequences can be used to identify homologues of the C. elegans
CC unc-53 gene. The UNC-53 protein can be used to identify proteins which
CC are active in the signal transduction pathway that can be used as
CC mentioned above.
XX
SQ Sequence 10288 BP; 2909 A; 2218 C; 2242 G; 2918 T; 1 other;

Query Match 4.8%; Score 72.2; DB 18; Length 10288;
Best Local Similarity 86.0%; Pred. No. 0.00054;
Matches 80; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

OY 1424 ctgtttatttttttaataaataaagatgaggaatttaaaaaaaaaa 1483
DB 8096 CTGTATTAFTTCTGTAAGAAAAAAGAAAAAAGAAAAAAGAAAAA 8037
OY 1484 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 1516

Db 8036 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 8004

RESULT 12

ID T71321/C

XX T71321 standard; DNA; 13414 BP.

XX T71321;

DF 11-SEP-1997 (first entry)

XX

DE Plasmid pCB50 encoding UNC-53 protein variant 7A used as bait vector.

XX

KW UNC-53; neuronal regeneration; revascularisation; wound healing;

KW neurodegenerative disease; Alzheimer's disease; Huntington's;

KW peripheral neuropathies; metastasis inhibition; cancer;

KW circular; Caenorhabditis elegans; ds.

XX

OS Synthetic.

XX

PN MO9638555-A2.

PD 05-DEC-1996.

XX

PF 31-MAY-1996; 96WO-EP02311.

XX

PR 31-MAY-1995; 95GB-0010944.

XX

PA (BOGA/) BOGAERT T.

PA (STRI/) STRINGHAM E.

PA (VAND/) VANDEKERCKHOVE J.

XX

PI Bogaert T, Stringham E, Vandekerckhove J;

DR WPI; 1997-034369/03.

XX

XX Caenorhabditis elegans UNC-53 protein 8A and 7A variants - useful to

PT promote neuronal regeneration, revascularisation or wound healing.

XX

PS Example 15; Page 132-139; 278pp; English.

XX

XX T71321 and T71322 are plasmids pCB50 and pCB51, respectively. They

CC are used as bait vectors in a yeast two hybrid system. pCB50 encodes

CC the full length UNC-53 protein variant 7A of Caenorhabditis elegans

CC and pCB51 encodes a C-terminal portion of this protein. Both vectors

CC were constructed by cloning the appropriate cDNA fragment into

CC pAS1-CYH2 (Clontech). The yeast two hybrid system was set up for

CC identifying proteins that interact with UNC-53. UNC-53 protein 8A and

CC 7A variants and nucleic acids encoding them are useful as medicaments

CC to promote neuronal regeneration, revascularisation or wound healing, or

CC for treatment of chronic neurodegenerative diseases (e.g. Alzheimer's or

CC Huntington's disease) or acute traumatic injuries. Transgenic cells and

CC organisms transfected with UNC-53 cDNA can be used to determine whether

CC a substance is an inhibitor or enhancer of the regulation of cell shape

CC or motility or the direction of cell migration by screening for a

CC phenotypic change in the cell. Inhibitors can be used to alleviate the

CC spread of disease inducing cells or metastasis. Probes derived from the

CC cDNA sequences can be used to identify homologues of the C. elegans

CC unc-53 gene. The UNC-53 protein can be used to identify proteins which

CC are active in the signal transduction pathway that can be used as

CC mentioned above.

XX

Sequence 13414 BP; 3854 A; 3072 C; 2884 G; 3603 T; 1 other:

SO

Query Match 4.8%; Score 72.2; DB 18; Length 13414;

Best Local Similarity 86.0%; Pred. No. 0.00056;

Matches 80; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

OY 1424 cgtgttatttttttaataaataagatgagagaatttaaaaaaaaaaaaaa 1483

DB 11222 CTGTATTATTCTATGAAAAAATAAAAAAAAAAAAAAAAAAAAAA 11163

OY 1484 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 1516

DB 11162 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 11130

RESULT 13

ID X37388

XX X37388 standard; cDNA; 773 BP.

XX

AC X37388;

XX

DT 06-JUL-1999 (first entry)

XX

DE Human secreted protein cDNA fragment containing gene 20.

XX

KW Human; secreted protein; prevention; treatment; protein therapy;

KW gene therapy; diagnosis; cancer; tumour; neurodegenerative disorder;

KW developmental abnormality; foetal deficiency; blood disorder; lymphoma;

KW leukemia; immune system disorder; autoimmune disease; hepatic disease;

KW renal disease; inflammation; allergy; asthma; sepsis; diabetes; AIDS;

KW Alzheimer's disease; cognitive disorder; schizophrenia; osteoporosis;

KW arthritis; psoriasis; digestive; endocrine; infection; ss.

XX

XX Homo sapiens.

XX

PN MO9909155-A1.

PD 25-FEB-1999.

XX

PF 18-AUG-1998; 98WO-US17044.

XX

PR 16-JUN-1998; 98US-0092956.

PR 15-JUL-1998; 98US-0092956.

PR 19-AUG-1997; 97US-0056368.

PR 19-AUG-1997; 97US-0056369.

PR 19-AUG-1997; 97US-0056535.

PR 19-AUG-1997; 97US-0056555.

PR 19-AUG-1997; 97US-0056556.

PR 19-AUG-1997; 97US-0056628.

PR 19-AUG-1997; 97US-0056629.

PR 19-AUG-1997; 97US-0056726.

PR 19-AUG-1997; 97US-0056728.

XX

PA (HMAN-) HUMAN GENOME SCI INC.

XX

PI Brewer LA, Duan R, Ebner R, Endress GA, Feng P;

PI Florence C, Florence KA, Komatsoulis GA, Lafleur DW;

PI Moore PA, Olsen HS, Rosen CA, Ruben SM, Shi Y, Soppet DR;

PI Young PE;

XX

DR WPI; 1999-190160/16.

DR P-PSDB; Y07763.

XX

XX New isolated human genes and the secreted polypeptides they encode

PT - useful for diagnosis and treatment of e.g. cancers, neurological

PT disorders, immune diseases, inflammation or blood disorders

XX

PS Claim 1a; Page 202; 280pp; English.

XX

XX This invention describes novel isolated human secreted proteins and

CC their encoding nucleic acid sequences. The products of the invention

CC are useful for preventing, treating or ameliorating medical conditions

CC e.g. by protein or gene therapy. Also pathological conditions can be

CC diagnosed by determining the presence or amount of expression of

CC the new polypeptides in a sample or by determining the presence or

CC absence of mutations in the new polynucleotides. Specific uses are

CC described for each of the 70 polynucleotides, based on which tissues they

CC are most highly expressed in, and include developing products for the

CC diagnosis or treatment of cancer, tumours, neurodegenerative

CC disorders, developmental abnormalities and foetal deficiencies, blood

CC disorders, leukemias, diseases of the immune system, autoimmune diseases,

CC hepatic and renal disease, lymphomas, inflammation, allergies, asthma,

CC sepsis, diabetes, Alzheimer's and cognitive disorders, schizophrenia,

THIS PAGE BLANK (USPTO)

Db 55 GACCGGGAGTTCCTCCGACCCCTGCCGCTTCCTCATCGTGGCCGAGATGCTTCTGGGG 114
Qy 245 ctgctggtggtgagcgtatgctgagcagcccgctacacacgtatccgagccatgctggtg 304
Db 115 CTGCTGTAATGAGACCTTATGCTGGAACGTGAGTACTTCCGGGCTCCCGCATTTGGCTGG 174
Qy 305 gtgatgttcgtcgtctctcctcgtcgtggtgacacgtcctcctcctacacacacgtc 364
Db 175 GTCATGTTTGTAGCTGTATTTTACTGGGTCCCTACCGTCTTCTTCATATATACATA 234
Qy 365 ttccagtcgacagaagttgtaacgttccctgagccacgtggtgttaatccttaac 424
Db 235 ACAATGACCTTACACAGCATTCCTCCGACCTGCCCTGGACAAAGTGGGCTGTGTAAAC 294
Qy 425 atcagcgcaccgtctctctacacacgcctcctacgcctcgtcgtcgtgagctgacgt 484
Db 295 GGCATGCTCTTGTCTTGTACTCTCTGC-----CGCTGTGTAGATGATCTTCCGCTC 348
Qy 485 acatccctgagggagcccgccctatacagcagcgagcgtcgtcgtcgtcgtcgtcgt 544
Db 349 TCCCTGAGAGGAGGACAGTACACTTCAACAGCTGGGGGCGCTCATCTTCTTCCCTTC 408
Qy 545 ttggtatgatacgtcctatgagtgagtgctcctcctcctacacacagcctggtgca 598
Db 409 CTGTCACATCTGCTACGCTGGAATATATTTCACTTTATAGCATGGAGA 462

RESULT 2
US-08-545-196B-10
; Sequence 10, Application US/08545196B
; Patent No. 6080577
; GENERAL INFORMATION:
; APPLICANT: MELKI, JUDITH
; APPLICANT: MUNNICH, ARNOLD
; TITLE OF INVENTION: SURVIVAL MOTOR NEURON (SMN) GENE: A GENE
; TITLE OF INVENTION: FOR SPINAL MUSCULAR ATROPHY
; NUMBER OF SEQUENCES: 65
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: BIRCH, STEWART, KOLASCH AND BIRCH, LLP
; STREET: PO BOX 747
; CITY: FALLS CHURCH
; STATE: VA
; COUNTRY: USA
; ZIP: 22040-0747
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/545,196B
; FILING DATE: 19-OCT-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: FARACI, C. J.
; REGISTRATION NUMBER: 32,350
; REFERENCE/DOCKET NUMBER: 2121-110P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 205-8000
; TELEFAX: (703) 205-8050
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1582 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; US-08-545-196B-10

Query Match 4.8%; Score 73.2; DB 3; Length 1582;
Best Local Similarity 86.2%; Pred. No. 6.5e-07;

Matches 81; Conservative 0; Mismatches 13; Indels 0; Gaps 0;
Qy 1423 actgtttatttttttaataaataaagatgaggaatttaaaaaaaaaaaaaaaaaa 1482
Db 1473 AATATTATTTTATTTTAAAAAATTTTAAAAAATTTTAAAAAATTTTAAAAAATTTT 1532
Qy 1483 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 1516
Db 1533 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1566

RESULT 3
US-08-545-196B-12
; Sequence 12, Application US/08545196B
; Patent No. 6080577
; GENERAL INFORMATION:
; APPLICANT: MELKI, JUDITH
; APPLICANT: MUNNICH, ARNOLD
; TITLE OF INVENTION: SURVIVAL MOTOR NEURON (SMN) GENE: A GENE
; TITLE OF INVENTION: FOR SPINAL MUSCULAR ATROPHY
; NUMBER OF SEQUENCES: 65
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: BIRCH, STEWART, KOLASCH AND BIRCH, LLP
; STREET: PO BOX 747
; CITY: FALLS CHURCH
; STATE: VA
; COUNTRY: USA
; ZIP: 22040-0747
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/545,196B
; FILING DATE: 19-OCT-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: FARACI, C. J.
; REGISTRATION NUMBER: 32,350
; REFERENCE/DOCKET NUMBER: 2121-110P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 205-8000
; TELEFAX: (703) 205-8050
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1582 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; US-08-545-196B-12

Query Match 4.8%; Score 73.2; DB 3; Length 1582;
Best Local Similarity 86.2%; Pred. No. 6.5e-07;
Matches 81; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

Qy 1423 actgtttatttttttaataaataaagatgaggaatttaaaaaaaaaaaaaaaaaa 1482
Db 1473 AATATTATTTTATTTTAAAAAATTTTAAAAAATTTTAAAAAATTTTAAAAAATTTT 1532
Qy 1483 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 1516
Db 1533 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1566

RESULT 4
PCT-US94-12883-3
; Sequence 3, Application PC/TUS9412883
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: UBIQUITOUS NUCLEAR RECEPTOR: COMPOSITIONS AND

TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-664-596B-3

Query Match 4.5%; Score 68.6; DB 1; Length 84;
Best Local Similarity 88.1%; Pred. No. 2.2e-06;
Matches 74; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 1432 tttttttaaaataaagatgaggaatttaaaaaaaaaaaaaaaaaaaaaa 1491
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1 TTTTNNAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAAA 60
QY 1492 aaaaaaaaaaaaaaaaaaaaaa 1515
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 61 AAAAAAAAAAAAAAAAAAAAAA 84

RESULT 10

US-08-738-367-3
; Sequence 3, Application US/08738367
; Patent No. 5827688
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
; APPLICANT: McCoy, John
; APPLICANT: Lavallie, Edward
; APPLICANT: Racle, Lisa
; APPLICANT: Merberg, David
; APPLICANT: Treacy, Maurice
; APPLICANT: Spaulding, Vilki
; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
; TITLE OF INVENTION: ENCODING THEM
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/738,367
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Brown, Scott A.
; REGISTRATION NUMBER: 32,724
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 498-8224
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 84 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-08-738-367-3

Query Match 4.5%; Score 68.6; DB 1; Length 84;
Best Local Similarity 88.1%; Pred. No. 2.2e-06;
Matches 74; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 1432 tttttttaaaataaagatgaggaatttaaaaaaaaaaaaaaaaaaaaaa 1491
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1 TTTTNNAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAAA 60

QY 1492 aaaaaaaaaaaaaaaaaaaaaa 1515
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 61 AAAAAAAAAAAAAAAAAAAAAA 84

RESULT 11
US-08-628-417-5
; Sequence 5, Application US/08628417
; Patent No. 5627054
; GENERAL INFORMATION:
; APPLICANT: GILLESPIE, DAVID
; TITLE OF INVENTION: COMPETITOR PRIMER ASYMMETRIC
; TITLE OF INVENTION: POLYMERASE CHAIN REACTION
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: U.S. ARMY CHEMICAL AND BIOLOGICAL
; ADDRESSEE: DEFENSE COMMAND
; STREET: OFFICE OF THE CHIEF COUNSEL (AMSCB-GC)
; CITY: ABERDEEN PROVING GROUND
; STATE: MARYLAND
; COUNTRY: USA
; ZIP: 21010-5423
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/628,417
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: BIFFONI, ULISES J
; REGISTRATION NUMBER: 39,908
; REFERENCE/DOCKET NUMBER: DAM 398-94
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 410-671-1158
; TELEFAX: 410-671-2534
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 140 bases
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: oligodeoxynucleotide
; HYPOTHETICAL: NO
; ANTI-SENSE: YES
; US-08-628-417-5

Query Match 4.5%; Score 68.4; DB 1; Length 140;
Best Local Similarity 83.0%; Pred. No. 2.9e-06;
Matches 78; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

QY 1423 acgtttattttttaaaataaagatgaggaatttaaaaaaaaaaaaaaaaaaaaaa 1482
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 20 ACTTAGAAATTAATTACTTAATAAANAAAAAANAAAAAANAAAAAANAAAAA 79
QY 1483 aaaaaaaaaaaaaaaaaaaaaa 1516
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 80 AAAAAAAAAAAAAAAAAAAAAA 113

RESULT 12
US-08-628-417-6
; Sequence 6, Application US/08628417
; Patent No. 5627054
; GENERAL INFORMATION:
; APPLICANT: GILLESPIE, DAVID
; TITLE OF INVENTION: COMPETITOR PRIMER ASYMMETRIC
; TITLE OF INVENTION: POLYMERASE CHAIN REACTION
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:

```

PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US9608736, 919
FILING DATE: 06-MAY-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Perkins, Patricia Anne
REGISTRATION NUMBER: 34,695
REFERENCE/DOCKET NUMBER: 2822-A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-587-0430
TELEFAX: 206-233-0644
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 1641 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 3..839
US-08-300-903A-8

Query Match      4.5% Score 68.2; DB 1: Length 1641;
Best Local Similarity 85.4%; Pred. No. 7.4e+06;
Matches 76; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 1428 ttatctttttaaataagaatgagagaattaaataaaaaaaaaaaaaaaa 1487
      |||||  |||||||  |||  |||  |||||  |||||  |||||  |||||
Db 1541 TCTATTTTCAMTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1600
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 1488 aaaaaaaaaaaaaaaaaaaaaaaa 1516
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 1601 AAAAAAAAAAAAAAAAAAAAAAA 1629

RESULT 14
US-09-014-969-14
Sequence 14, Application US/09014969
Patent No. 5965397
GENERAL INFORMATION:
APPLICANT: Jacoby, Kenneth
APPLICANT: McCoy, John M.
APPLICANT: Lavallee, Edward R.
APPLICANT: Racie, Lisa A.
APPLICANT: Merberg, David
APPLICANT: Treacy, Maurice
APPLICANT: Spaulding, Vikki
APPLICANT: Agostino, Michael J.
TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
TITLE OF INVENTION: ENCODING THEM
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: MA
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/014,969
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Sprunger, Suzanne A.
```



```

1      REGISTRATION NUMBER: 41,323
2      TELECOMMUNICATION INFORMATION: none
3      TELEPHONE: (617) 498-8284
4      TELEFAX: (617) 876-5851
5      INFORMATION FOR SEQ ID NO: 14:
6      SEQUENCE CHARACTERISTICS:
7      LENGTH: 2447 base pairs
8      TYPE: nucleic acid
9      STRANDEDNESS: double
10     TOPOLOGY: linear
11     MOLECULE TYPE: CDNA
12     US-09-014,969-14

```

Query Match	4.5%	Score 67.8;	DB 2;	Length 2447;
Best Local Similarity	80.4%;	Pred. No. 1e-05;		
Matches	78;	Conservative	1;	Mismatches 0;
				Gaps 0

<i>Oy</i>	1420 gtgcgtcttactatcttttcacaaacaagaagtaggaagtcaaaaaaaa	1479 : :
Db	2222 GTGCACATTAACCTTTTGGGCAAAAAAAAAAARAAAAAA	2281 : :
<i>Oy</i>	1480 aaaaaaaaaaaaaaaaaaaaaaa 1516	
Db	2282 AAAAAAAAAAAAAAAAAAAAAAAA 2318	

RESULT 15
US-08-897-340-4

```

: GENERAL INFORMATION:
: APPLICANT: Gliseno, Carlos J. and Errada, Patrick, R.
: TITLE OF INVENTION: Weight Control Pathway Genes and Uses
: TITLE OF INVENTION: Therefor
: NUMBER OF SEQUENCES: 36
: CORRESPONDENCE ADDRESS:

```

```

? COMPUTER READABLE FORM.:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: PatentIn Release #1.0, Version #1.25
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/897,340

```

Query Match 4.48; Score 67.4; DB 2; Length 1700,

Best Local Similarity 63.38; Pred. No. 1.1e-05;
Matches 95; Conservative 4; Mismatches 51; Indels 0; Gaps 0.

QY 1367 caaagggagggaagcaagcgcttcgcttggagatctcaatggttgggtgggactg 1420

Db 1508 CTAGGGGAGGATGGAACCCNCGGTAKACAGGAACACAGCGCTTGCAATTTTGTGG 1567

QY 1427 ttctattttttcatataaataagatgatgagaaattaaaaaataaaaaaataaa 1480

Db 1568 TATTGTTAACTTTATTATTTAAAAAGAAAACAACTCTGTAATTTTAAAAAGAAAAGAT 1627

QY 1487 aaaaaaaaaaaaaaaaaaaaaaaaaa 1516

Db 1628 TAAAAAAAAAAAAAAAAAAAAAAAAA 1657

```
Search completed: March 1, 2001, 02:07:01
Job time: 12792 sec
```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 28, 2001, 19:43:14 ; Search time 8425.79 seconds

(without alignments)
1731.065 Million cell updates/sec

Title: US-09-602-597-1

Perfect score: 2850

Sequence: 1 aagccctgaaggggtcacaag.....aaaaaaaaaaaaaaaaaaaaa 2850

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1118133 seqs, 2558875100 residues

Total number of hits satisfying chosen parameters: 2236266

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:*
1: gb_bal:*
2: gb_baz:*
3: gb_om:*
4: gb_ov:*
5: gb_ph:*
6: gb_pl1:*
7: gb_pl2:*
8: gb_pl3:*
9: gb_pr2:*
10: gb_pr3:*
11: gb_ro:*
12: gb_sy:*
13: gb_un:*
14: em_fun:*
15: em_hum1:*
16: em_hum2:*
17: em_in:*
18: em_om:*
19: em_or:*
20: em_ov:*
21: em_pat:*
22: em_ph:*
23: em_pl:*
24: em_ro:*
25: em_sts:*
26: em_sy:*
27: em_un:*
28: em_vl:*
29: gb_htg1:*
30: gb_htg2:*
31: gb_in1:*
32: gb_in2:*
33: em_bal:*
34: em_baz:*
35: em_hum3:*
36: em_hum4:*
37: gb_pr4:*
38: gb_htg3:*
39: gb_htg4:*
40: gb_htg5:*
41: gb_htg6:*
42: gb_htg7:*
43: em_htg1:*

44: em_htg2:*
45: em_htg3:*
46: em_hum5:*
47: gb_pl3:*
48: gb_pr5:*
49: gb_htg8:*
50: gb_htg9:*
51: gb_htg10:*
52: gb_htg11:*
53: gb_htg12:*
54: gb_htg13:*
55: gb_htg14:*
56: gb_in3:*
57: gb_htg15:*
58: gb_htg16:*
59: gb_htg17:*
60: em_htg4:*
61: em_htg5:*
62: em_htg6:*
63: em_htg7:*
64: em_hum6:*
65: gb_htg18:*
66: gb_htg19:*
67: gb_htg20:*
68: gb_htg21:*
69: gb_htg22:*
70: gb_htg23:*
71: gb_vil:*
72: gb_vl2:*
73: gb_ba3:*
74: em_htg8:*
75: em_htg9:*
76: em_htg10:*
77: gb_pr6:*
78: gb_pr7:*
79: gb_sts1:*
80: gb_sts2:*
81: gb_pat1:*
82: gb_pat2:*
83: em_htg0:*
84: gb_htg24:*
85: gb_pr8:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2029	71.2	221262	30	AC009090 Homo sapi
2	2029	71.2	237448	51	AC023825 Homo sapi
3	1941	68.1	1990	37	AK025452 Homo sapi
4	1900.4	66.7	2159	48	AX017393 Sequence
5	1818.8	63.8	2011	48	AX017306 Sequence
6	1505	52.8	2020	37	AK023291 Homo sapi
7	198.8	7.0	221262	30	AC009090 Homo sapi
8	191.4	6.7	33077	44	AC012092 Homo sapi
9	92	3.2	7218	81	166494 Sequence 14
10	82	2.9	201341	69	AL391825 Homo sapi
11	80.4	2.8	162093	53	AC027070 Homo sapi
12	74.2	2.6	177816	41	AC017643 Drosophila
13	74.2	2.6	262395	31	AE003802 Drosophila
14	74	2.6	129891	29	AC007697 Drosophila
15	73.4	2.6	3327	78	HSMB00260 Homo sapi
16	71.6	2.5	2773	78	HSMB00446 Homo sapi
17	71.4	2.5	3482	78	HSMB00550 Homo sapi
18	71	2.5	705	78	HSMB00237 Homo sapi
19	70	2.5	3787	78	HSMB01412 Homo sapi
20	69.6	2.4	363	32	AF135821 Mesobuthus
21	69.6	2.4	3386	78	HSMB00161 Homo sapi

```
22 69.4 2.4 1388 37 AK027096 Homo sapi
23 69.2 2.4 788 79 CNS01JAM
24 69 2.4 987 10 AF116698 Homo sapi
25 69 2.4 1022 47 PSP1A45R
26 68.8 2.4 665 78 HSM800251
27 68.8 2.4 1977 11 AF079763
28 68.8 2.4 9711 71 AF177036
29 68.8 2.4 237448 51 AC023825
30 68.6 2.4 1285 78 HSM801947
31 68.6 2.4 1438 78 HSM800279
32 68.4 2.4 1851 37 AK026744
33 68.4 2.4 2041 78 HSM802525
34 68.4 2.4 2571 10 AF113676
35 68.2 2.4 446 32 AF146743
36 68.2 2.4 1252 78 HSS7A
37 68.2 2.4 2550 37 AK025772
38 68.2 2.4 194180 58 AC069140
39 68 2.4 1686 21 E08995
40 67.8 2.4 2462 78 HSM800419
41 67.8 2.4 2972 10 AF004162
42 67.8 2.4 15528 12 PEAVGEN
43 67.8 2.4 15528 81 A93016
44 67.6 2.4 1307 81 AR083266
45 67.6 2.4 1824 10 AF111851
```

ALIGNMENTS

RESULT 1

AC009090

AC009090 221262 bp DNA HMG 21-JUN-2000
Homo sapiens chromosome 16 clone RP11-407G23, WORKING DRAFT

SEQUENCE, 26 unordered pieces.

AC009090 7 GI:8575954

HMG; HTGS_PHASE1; HTGS_DRAFT.

human.

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 221262)

DOE Joint Genome Institute.

Unpublished

2 (bases 1 to 221262)

DOE Joint Genome Institute.

Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

On Jun 21, 2000 this sequence version replaced gi:7689913.

-----Genome Center

Center: Joint Genome Institute

Center Code: JGI

Web site: <http://www.jgi.doe.gov>

Project Information
Center Project Name: 562820

Center clone name: RPCI-11_407G23

Summary Statistics

Consensus quality: 188611 bases at least Q40

Consensus quality: 204128 bases at least Q30

Consensus quality: 207087 bases at least Q20

Estimated insert size: 218870; agarose-fp estimation

Estimated insert size: 218762; sum-of-ctg's estimation

Quality coverage: 6.45 in Q20 bases; agarose-fp estimation

Quality coverage: 6.45 in Q20 bases; sum-of-ctg's estimation.

* NOTE: This is a 'working draft' sequence. It currently

* consists of 26 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```
1 1225: contig of 1225 bp in length
1 1325: gap of unknown length
1 1326: contig of 1222 bp in length
1 2547: gap of unknown length
1 2648: gap of 1451 bp in length
1 4098: contig of 1451 bp in length
1 4198: gap of unknown length
1 4199: contig of 1996 bp in length
1 6185: gap of unknown length
1 6295: gap of 1104 bp in length
1 7398: gap of unknown length
1 7499: gap of 1392 bp in length
1 8890: gap of unknown length
1 8991: contig of 1084 bp in length
1 10074: contig of 1084 bp in length
1 10175: gap of unknown length
1 10175: contig of 1248 bp in length
1 11422: gap of unknown length
1 11423: contig of 1510 bp in length
1 11523: gap of unknown length
1 13033: contig of 1110 bp in length
1 13133: gap of unknown length
1 14243: gap of unknown length
1 14343: gap of 1205 bp in length
1 15348: gap of unknown length
1 15647: gap of 1639 bp in length
1 17287: gap of unknown length
1 17386: gap of 1583 bp in length
1 18969: contig of 1583 bp in length
1 18970: gap of unknown length
1 19070: contig of 2365 bp in length
1 21435: gap of unknown length
1 21435: contig of 1921 bp in length
1 23455: gap of unknown length
1 23456: gap of 1228 bp in length
1 24784: gap of unknown length
1 24884: gap of 3003 bp in length
1 27887: gap of unknown length
1 27986: gap of 2736 bp in length
1 30722: contig of 2736 bp in length
1 30723: gap of unknown length
1 30822: gap of 4380 bp in length
1 30823: contig of 4380 bp in length
1 35203: gap of unknown length
1 35203: gap of 10463 bp in length
1 35303: gap of unknown length
1 45766: gap of unknown length
1 45865: gap of 15054 bp in length
1 60919: contig of 15054 bp in length
1 60920: gap of unknown length
1 61020: contig of 17000 bp in length
1 78020: gap of unknown length
1 78120: gap of 23944 bp in length
1 102064: contig of 23944 bp in length
1 102163: gap of unknown length
1 102164: contig of 22065 bp in length
1 124229: gap of unknown length
1 124328: gap of 37191 bp in length
1 124329: contig of 37191 bp in length
1 161520: gap of unknown length
1 161520: contig of 59643 bp in length.
```

FEATURES

source

1. 221262

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="16"

/clone="RP11-407G23"

/clone_lib="RPCI human BAC library 11"

BASE COUNT 57404 a 53185 c 51551 g 56602 t 2520 others

ORIGIN

Query Match

Best local Similarity 99.7%;

Matches 2043; Conservative 0; Mismatches 5; Indels 1; Gaps 1;

DB 169565 ATTCGACAGCCCTCATCTCGAATCTCTCGGAACACCTCCGATGCGCCCTCATC 169624

OY	814	caatgcccctctgctgcgaagtatgtatgcgaatcctccaaagccctgggtgctactctggg	873
Db	169625	CACGCCCCCTCTGCTGCGAGTATGATGCGCATCTCCAGGCCCTGGGTGCTACTCTGGG	169684
OY	874	agcaagcgactccgagttccagctcaagaagcggaagcaacatcaatgacacccggaagatt	933
Db	169665	AGCAACGACTCCGAGTCCAGCTCAGACAGCGAAGGACATCAATGCGACCGGAAAGATT	169744
OY	934	gtctctccatcttcctgcgaaccaaaccttccctcgaagcccccctagttcttcgltccctac	993
Db	169745	GTCTCTCCATCTTCGCCAACAACACCTTCCTCGAGGCCCCCTGATTCTTCGCCCTTAC	169804
OY	994	acaaggagctctcccccgaaggtagatcgcgaacgttcaatgctgcctatagcatatgtc	1053
Db	169805	ACAGGAGACTCTCTCCCCAAGGGTAGATCGGACCCCTTCATGCTCGCTTATAGCATTTATGTC	169864
OY	1054	cctcaaaaaaaactcccttggccgtacatccgtgttaacaatatgacattttaaccaatcc	1113
Db	169865	CCTCAAAAAAAACCTCTTTCCTGCGCATCTGTTACAACTGACATTTTAAACCAATTC	169924
OY	1114	aatcctaaaatgtcccaaaatccaaacactgtgccccgaatcggtgttggttccctcttctac	1173
Db	169925	AATCTAAAAATGTGCGCAAAATCCACACTGTGTGGCCGATTCGTGTGGTTCTCTTCTTAC	169984
OY	1174	tcactgcagatagcaaaacctgtcccgctgtgcaacttccctactgatalttggagagag	1233
Db	169985	TCCACTGCGAGTATGACCAAAACCTGTCCGCTGCCACTTTCCTACATGATTTTGGAGGAGG	170044
OY	1234	gcaagggcccaagcgaagtctgacactaaaatgccccgaagaatatggaccggcgctgttg	1293
Db	170045	GCAAAGCCCCACCGAAGTCCACTAAAAATGCCCCAGAGAAATGAGCAACGGCGGCTTG	170104
OY	1294	ccaaaggtcttgaggtttatatgctcttcgttttcttcttccctccgcagacaaagaagta	1353
Db	170105	CCAAAGGTTTGGGTTTATGTGCTTCGTGTTTTCCTTTCGCCAGCAACAAAGAGTA	170164
OY	1334	agggcagttataggacaaggttatattaacatctcatgttaaatgaaatgtgtgtttgg	1413
Db	170165	AGGGCAGTTATTTGGACAGAGTGTATTTAAACATTCTATTGAAATGATGTGTGTTGG	170224
OY	1414	ttctactgcatttggagcatgctgggggaagaagacgcaccggtatgaaatggagcc	1473
Db	170225	TTCTACTCATTTGGAGCAATGCGGGGGAAGAAATGACCCAGTATGTAATAATGGAGCC	170284
OY	1474	cttcccttgaaactaacacagctcctctgatatgtgtgtaactaagtaaaga tgaatacccat	1533
Db	170285	CTTCCCTGGAACTAACCAAGTCTCTGTGATTTGTGTGACTAAGTAAGAATGAATAAACCCAT	170344
OY	1534	ctgctggygggtgtcaactcaactccgcatgtcatgtgtaaagccttccatacccttggcc	1593
Db	170345	CTGCTGGGGGGGTCACTTCACACTTCGGCATTTGTGAAAGCTTTCATACCCCTTGCC	170404
OY	1594	attccctctctccctctctctccaaaccccatttaagcaaggaaagagactgctcaaaagaag	1653
Db	170405	ATTCCCTCTCTCTCTCTCTCCAAACCCCATTTATGACGGAAGGACATGCTAACAGAGAG	170464
OY	1654	cttccatctcaaaccttctctctgctcgtgggaatatattatgtgttttttgaataaa	1713
Db	170465	CTTCCATCTCAAAACCTTTCCTCTGCGCGGGAATTTATTTATATGTTTGTTTTGAAATAAA	170524
OY	1714	ggatttagtttaagattcctaataatttagagaacaacacgttagccctgttactaataagc	1773
Db	170525	GGATTTAGTTTAAATCTTAATTTTATGAGAAACAAACGTAAGGCGCTTGTACTTAATAGC	170584
OY	1774	cagacatcaagacgcgcaggtaggtatgtttaatgtagatctatttcttgcaagctcctgg	1833
Db	170585	CAGACATCAAGACGCAAGTAGATGTATTAATGAGATGACTTATTTCCTGGACGCTCTGG	170644
OY	1834	aatccctaatctgttaaatgagtgtagacaacacttcatatgttgacaattctatattgagcc	1893
Db	170645	AATCCTAATATTGTAAATGATGGGACACACTTCGATATTGTGACCAATTCTATTGAGGCC	170704
OY	1894	c-tctcgtttaaagtcatattatactgtggtctttaaactgtlgaaatctlatittcctaacta	1952

Db	170705	CTTCTCTGTTTAAGCATATTATACCTGTGTGCTTTTAACGTGGAAATCTATTCTTAACCTTA	170764
QY	1953	aagggcgcgccttagtactttctctgtgcctctgcgcctcttttccctccaagaag	2012
Db	170765	AAGGAGCGCCCTACTACTTTTCTTTTGGTGGCTCGCTCGCTCTTTTCTTTCCAAACAG	170822
QY	2013	caactctgaggccatgagcaagccaanaactagaggtagctgtccacctcgtctcataaag	2072
Db	170825	CAACTCTGAGGCGCATGAGCAGCCAAAACCTAGAGGTACTGCTCCACCGTCTCATFAAG	170884
QY	2073	ggaaaagggtccatccctcttgatctctggaaggagagaggaagtgtgtggaagcctcg	2132
Db	170885	GGAAAAGGGGCTATCCCTTGGATTCTGGAGAGGAGAGGAGAGATGTGTGGAGGCTTCG	170944
QY	2133	aggaaagagatagacatgagtttgacaanaactgtgaggtccctgcgtcttaagaatag	2192
Db	170945	AGGAGAGAGATAGACATGAGCTTTGACACACATCTGTAGGGCTCTCTGCTTTAGAAATAG	171004
QY	2193	catgtaccatctttatccatctccctctatccctctacatacaatctgttttacttctctggg	2252
Db	171005	CATGTACCATTTCTTTATTCATTTCCCTTTATTCCTATTCATACATATGTTTTTACTTTCTTGGG	171066
QY	2253	tgtgagactgtgtgagacacacacaaaatgtgttgacaactgtgtatgccggcagcagaagc	2312
Db	171065	TGTGTGACTGTGATGAGACACACAAAATGTGTGTGACACTGTGATGTGGCGGAGCAGCAAGC	171124
QY	2313	agctactgactcttgaaacttgagagagagagagccctgtgatctccatccaaagccaccccttt	2372
Db	171125	AGCTACTACTTTTGAACATGGGGCAGAGAGGCCCTCGATCTCATTCACGCCACATCTCTTTT	171186
QY	2373	ccccttccagtagcagtgacactctgtgtgcccattgtgcagatgagcagcttccctgcacca	2432
Db	171185	CCCCCTTCCAGTAAAGATGACACTGTGTGCCCATTTGGCAGATGGCGACTTCCCTGCACCCA	171244
QY	2433	taactgtagcttctgtgaattcttccctcttccagaaactactctgtgtctaattgtctgc	2492
Db	171245	TAACTGATGCTTTGTGAATTTCTTCCCTTTTCAGAACTACTCTGTGCTAATTTGTTCTGC	171304
QY	2493	cagttatgggcacatcagctcactcctgaacaaagaagcatttaagtaaaactctgtatggc	2552
Db	171305	CAGTATGGGGCGCATCAGTCCATCTGTACAAACAAAGACATTTAGGTAAACCTTTGTATGGC	171364
QY	2553	acctctgtctctctgtctcatctgttccctgtgatagtcctgtgttataccaagatgtac	2612
Db	171365	ACCTCTGCTCTTCTGTGCTTACTTTGTTCCTGTGATAGTCCGTGTATTATACAGCATGTAC	171424
QY	2613	ccaaaacagcctcacatgtgttacagagagcagccagaacatcaaaatcatcatcttatt	2672
Db	171425	CCAAAACAGCCCTCACATTTGTACAGAGGCGAGCGCAGACATCAAAATCATCATCTTTAT	171484
QY	2673	gtgagatgactccttaagaagcgcattactgtrtccatcagagcctcttggttggaaagaagt	2732
Db	171485	GTGGGATGACTCTTTAAGGGCCATTACTCGTATTCATATGCGCTCTTGTATGTGGAAAGAGT	171544
QY	2733	tgacagagaggttcagaggttttaaaaacatccattacatcgaagaagctataaaacctgtcag	2792
Db	171545	TGACAGAGAGGTTTGACGGGTTTAAAAACATCCATTAACTGAAAGCTATATAAAGCTGTACG	171604
QY	2793	agaacaanaa 2801	
Db	171605	AGAACACAGA 171613	
RESULT 2			
AC023825			
LOCUS	AC023825	237448 bp	DNA
DEFINITION	Homo sapiens chromosome 16 clone RP11-322D14, WORKING DRAFT		03-SEP-2000
ACCESSION	AC023825		
VERSION	AC023825.5	GI:9965535	
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT.		
SOURCE	human.		

D	74331	CTGCGGGGGTCACTTCACTCGGATGCAATTGTGAACCTTTCCATTACCTTTGGCC	74330
Q	1594	atccctctctcctctctcctccacccttattatgcaggaagagctctcaacaagac	1653
D	74391	ATTCCCTCTCTCTCTCTCCAMACCAATTTATCGAGAAAGGAGCTGTAAACAAGACG	74450
Q	1654	ctccatctcaaaccttctcctgcctgggaattattatgattgttttgaataaa	1713
D	74451	CTTCCATCTCAAACTTTTCTCTGCTGGGAATTTATTTATGTTTGTTTTGAATAA	74510
Q	1714	ggatttagttaagaattcctaattttaagaacaacagtaagccttcttactaatagc	1773
D	74511	GGATTAGTTTAAGATTCTTAATTTTAGAGAAACAAGAGTAGCCCTGTTTACTAATAGC	74570
Q	1774	cagacatacgaacctgcaggtagtagtgttaatagatgactattcttcgcagctctgg	1833
D	74571	CAGACATCAAGAACTCCAGGTAGGTATGTTAAGACATCACTTAATTTCTGGCAGCTCCGG	74630
Q	1834	aatccataattgttaaatgaagtggagcaacttgatattgagaccattctattagcc	1893
D	74631	AATCCTAATATTGTAAATGAGGGGACACACTGCAATTTGTGACATTCATTAAGGCC	74690
Q	1894	c-tccctgtttaatgcataattactgtgtctttaaactgtgaactcaattctaacca	1952
D	74691	CTTCTCTGTTTAATGCAATATTACTTGGCATTAACTGTGAATCTATTCTTAACCTA	74750
Q	1953	aaggtgcgcctagtaacttcttctgcctcctgcctgcctcttctccccaagaag	2012
D	74751	AAGGTCGCGCCCTAATTACTTTTCTTGGCGCTCGTGCCTCTCTTTCCTTTCCAAACAG	74810
Q	2013	caaccttgaggccatggaagcagccaaaacttgaggtactgctccacctcgttctaagaag	2072
D	74811	CAACCTGAGGGCAAGACAGCAAGCAAAACTAAGGTACTGCTCCACCTCGTCTCAATAAG	74870
Q	2073	ggaacggggcatccctccttgaattcttgagagagagagagagatgagtgttgaagcccg	2132
D	74871	GGAAACGGGGCTATCCCTTGGAATTCTTGAGAGAGGAGAGGAGATGCTGTGAAGCCTCG	74930
Q	2133	aggacagatatagatagacttctgacaacaactctgtagctctcctcctttagataag	2192
D	74931	AGGACAGAGATTAAGACATGAGCTTTGACAACAATCTGTGAGGCTCTCTGCTTAAATAAG	74990
Q	2193	catgtaccattcttattccattccctctattcctctacatcaattgttttacttcttggg	2252
D	74991	CATGTACCAATCTTATCCATTCCCTTATTCCTATCCATCAATGTGTTTACTTCTTGGG	75050
Q	2253	tgtgagacttgatgagacacacacaaaatgtgtgaactgtatgacggcagagagagc	2312
D	75051	TGTGAGACTGTAGTAGACACACAAAATGTGTACACTGTGATCCGGCAGCAGAGAC	75110
Q	2313	agctactgacttggacatgggcagagagccctctgatactcatccagcccectcctt	2372
D	75111	AGCTACTGACTTTGAAACATGGGACAGAGAGGCCCTGTGATTCATCCAGCCACTCTTTT	75170
Q	2373	ccccctcagtaacagtgacaactctgggtgccatctggcagatvgggaactcctgcacca	2432
D	75171	CCCCCTCAAGTAACAGTAGACACTCTGGTCCCACTTGGCAGATGGCACTTCCTGCACCCA	75230
Q	2433	taacagatgacttggaaattcctcctcttccctttcagaactactctgtgtaattgtctgc	2492
D	75231	TAACTAGTGCTTTTGAAATTTCTTCTCTCTTTCATCAAACTACTCTGTGCTAAATGTCTGC	75290
Q	2493	cagta.tgggcgcataagctccatccctgacaaacaagacatttaagtaaaactcttgaagc	2552
D	75291	CAGTATGGGCGGATGAGTCCATCCTTGAAACAAACAGACATTTAGGTAAACCTTTGTAGGC	75350
Q	2553	acctctgcctctcctgctcaatgtctcctgtgatagtccctgttgttatacagcatgtac	2612
D	75351	ACCTTCTGCTCTCTGCTTCAATTTGCTGTGATGTGCTGTGTTATTACAGCATGTAC	75410
Q	2613	ccaaaagcctcacatggtttacaggaagagccagagacatcaaaagcatcatctttat	2672
D	75411	CCAAAAGCCTCACATTTGTTACAGAGAGCGCCAGACATCAAAAGTCATCATCTTTAT	75470

QY	2673	gtggatctacctttaagagccattacacgtatcctcatgctgcgccttgatgtggaagaagt	2732
Db	75471	GTGGATCTACCTTTTAAGAGCCATTACTGTATCTCATGGCCCTTGATGTGGAAACAAGT	75530
QY	2733	tgacagaggttctcagggtttaaaaacatccattacaatgtaagcctaataaactgtcag	2792
Db	75531	TGACAGAGGTTTGACAGGCTTTTAAAAACATCATTAATGAAGTAAGCTAATAAACCTGTGAG	75590
QY	2793	agaacaaa 2801	
Db	75591	AGAACAAA 75599	
RESULT	3		
AK025452			
LOCUS	AK025452	1990 bp	29-SEP-2000
DEFINITION	Homo sapiens cDNA: FLJ21799 fis, clone HEP00611.		
ACCESSION	AK025452		
VERSION	AK025452.1	GI:10437969	
KEYWORDS	oligo capping; fis (full insert sequence).		
SOURCE	Homo sapiens: hepatoma cell_line:hepg2 cDNA to mRNA, clone_lib:HEP clone:HEP00611.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
AUTHORS	Kawabata,A., Hiki,T., Kobatake,N., Inagaki,H., Ikema,Y., Okamoto,S., Okitani,R., Ota,T., Suzuki,T., Obayashi,M., Nishi,T., Shibahara,T., Tanaka,T., Nakamura,Y., Isogai,T. and Sugano,S.		
TITLE	NEDO human cDNA sequencing project		
JOURNAL	Unpublished (2000)		
AUTHORS	2 (bases 1 to 1990)		
TITLE	Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T., Shibahara,T., Tanaka,T. and Nakamura,Y.		
JOURNAL	Shibahara,T., Tanaka,T. and Nakamura,Y.		
COMMENT	Direct Submission Submitted (29-AUG-2000) to the DDBJ/EMBL/GenBank databases. Sumio Sugano, Institute of Medical Science, University of Tokyo, Laboratory of Genome Structure Analysis, Human Genome Center, Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan (E-mail:chna@ims.u-tokyo.ac.jp, Tel:81-3-5449-5286, Fax:81-3-5449-5416) NEDO human cDNA sequencing project supported by Ministry of International Trade and Industry for Biotechnology: cDNA library sequencing: Research Association for Biotechnology: Department of Virology and Human Genome Center, Institute of Medical Science, University of Tokyo (partly supported by Science and Technology Agency).		
FEATURES	Location/Qualifiers		
source	1..1990		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/cell_line="HepG2"		
	/cell_type="hepatoma"		
	/clone="HEP00611"		
	/clone_lib="HEP"		
	/note="cloning vector pME18FL3"		
	235..999		
	/note="unnamed protein product"		
	/codon_start=1		
	/protein_id="BAB15136.1"		
	/db_xref="GI:10437970"		
	/translation="MDGSDGDNILIIKKFVSEAFIDERRRRKROEEMEKVRKPEDPECCPEEYVDRSLYERKIQEOKDQOEYEDQEFKKNYKRGIDEDETNFDLDEYSNOGLIEKQRREELKELEKYENNNLKAVISQENKREVKRLVFKIETKKNFSQAKLLAGAVKRRSSSGNSVYKRLKPDPEDDKQNPSSCSLGNSTLSGSPSHCPAAVACITLIPSLGAMSGSSDSSSSDSECTINATGKIVSISFETNFTLEAP"		
CDS	562 a 447 c 459 g 522 t		
BASE COUNT			
ORIGIN			
Query Match	68.1%	Score:1941; DB 37; Length:1990;	

Db	1992	16CTCACCCTGCTTCATTAAGGAACGGGCTCATTCCTTGATTCCTGAGAGGAGAA	2051
Qy	2111	ggagatggtgtgtgagggcctcgagacagagatagacatgagctttagacaacatcttga	2170
Db	2052	GGGAGATGGGTGTGGAGG-CTCGAGGACAGACAGACATGAG-TTGACAACAATCTGTA	2109
Qy	2171	ggcttcctgctttagaataagca	2194
Db	2110	GGC-GCCTGGCTTAGAATTAAGCA	2132
RESULT	5		
AX017306	AX017306	2011 bp	DNA
LOCUS	AX017306	Sequence 62 from Patent WO947669.	PRI
DEFINITION	AX017306		07-SEP-2000
ACCESSION	AX017306		
VERSION	AX017306.1	GI:10042224	
KEYWORDS			
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
	1 (bases 1 to 2011)		
TITLE	Schmitt,A., Spectt,T., Dahl,E., Hinzmann,B., Rosenthal,A. and		
JOURNAL	Pilarsky,C.		
	Human nucleic acid sequences from tissue of breast tumors		
	Patent: WO 9947669-A 23-SEP-1999;		
	SCHMITT ARMIN (DE); SPECTT THOMAS (DE); DAHL EDGAR (DE); HINZMANN		
	BERND (DE); ROSENTHAL ANDRE (DE); METZGEN GBS FUER GENOMORSCHEUN		
	(DE); PILARSKY CHRISTIAN (DE)		
FEATURES	Location/Qualifiers		
Source	1..2011		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
BASE COUNT	573 a 467 c 448 g 523 t		
ORIGIN			
Query Match	63.8%; Score 1818.8; DB 48; Length 2011;		
Best Local Similarity	99.4%; Pred. No. 0;		
Matches 1868; Conservative	0; Mismatches 7; Indels 5; Gaps		
Qy	195	gtaggttggttcttcattatgtagtggagggagtagttagttaaactattatcaaaaagag	254
Db	133	GCAGATGCTGTTTCATTATGATGAGGCGGATGATGTTAACCCTTATATCAAAAAGAG	192
Qy	255	gtttgtgctgagagagactagatgaacggcgcaaaagagagcaagaagtggagaga	314
Db	193	GTTTGTGCTGAGGAGAACATGATGAACGGCGCAAAAGAGCAAGAAATGGAGAGA	252
Qy	315	agttgaaaacctgagaagatccagaagaatgtccagagaggtttagacctcgatctct	374
Db	253	AGTTGAAAACCTGGAAGATCCGAAAGANTGTCCAGAGAGGTTTATGACCCCTGATCTCT	312
Qy	375	atataaaggtctacaggaacagaagaagcaagaagcagcaagagtaacgagaaacagtlcaa	434
Db	313	ATATGAAAAGGCTACAGGAACAGAAAGACAGAGACAGACAGAGTACGAGAGAACAGTTCAA	372
Qy	435	attcaaaaacatgtaagaaggtctgatagaagatagagaccacttcccttgatgaagtttc	494
Db	373	ATTCAAAAACATGATGAAGGCTGTGATGAAATGAGACCACTTCTTGATGAGGTTTC	432
Qy	495	tcgacagcaggaactaatagaaaaaacagcaagaagaagaagaagactgaagaagta	554
Db	433	TCGACAGCAGGAACATATATRAGAAAACCAACGAAGAGAGAGAACTGAAAAGACTGAAGA	492
Qy	555	atacagaataaactccaagaaggttggaattctcaagaagaaacaagaagaagttgnaaa	614
Db	493	ATACGAATAATACCTCAGAAAGTTGGAAATTTCTCAAGAGAACAGAGAGAGTGGAAAA	552
Qy	615	gaactgactgtgaagcctatagaaacaaagaagaagttctccagggcgaagctgtggc	674

[illegible]

QY	1755	gagcctgttacttaabagccagacatcgaactcgcagtgaggtatgttbaatgaagact	1814
Db	1692	ggccttgtttacttaatagccacagacatcagacttgcagtgaggtatgttbaatgaagact	1751
QY	1815	tattctggcagctccttggaaatcctaataitgtlaaatgagtggagacacacttgcataatg	1874
Db	1752	tatttttggcagactccttggaaatcctaataatgttaaatgagtgaggacacacttgcataatg	1811
QY	1875	tgaccattctatgtggccc-tctctgtttaatgtcatataacttgbtctttaaactgt	1933
Db	1812	tgaccattctatgtggcccctctctgtttaatgtcatataacttgbtctttaaactgt	1871
QY	1934	ggaactctattcttaaccctaaagtgctgcgccagctatttc-tttgcctcctgcctgc	1992
Db	1872	ggaactctattcttaaccctaaagtgctgcgccagctatttc-tttgcctcctgcctgc	1931
QY	1993	tcttttccctt--ccaacagcaactctgagggcatgagcagccaaactagaggtac	2050
Db	1932	tcttttccctt--ccaacagcaactctgagggcatgagcagccaaactagaggtac	1991
QY	2051	tgctccacctgtctcctacaa 2070	
Db	1992	tgctccacctgtctcctacaa 2011	
RESULT	6		
LOCUS	AK023291	2020 bp	29-SEP-2000
DEFINITION	Homo sapiens CDNA FLJ13229 f1s, clone OVARC1000106.		
ACCESSION	AK023291		
VERSION	AK023291.1 GI:10435163		
KEYWORDS	oligo capping, f1s (full insert sequence).		
SOURCE	Homo sapiens ovary, tumor tissue CDNA to mRNA, clone-lib:OVARC1 clone:OVARC1000106.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
AUTHORS	1 (sites)		
	Isogai,T., Ota,T., Haysshi,K., Sugiyama,T., Otsuki,T., Suzuki,Y., Nishikawa,T., Nagai,K., Sugano,S., Ishibashi,T., Fujimori,K., Tanai,H., Kinata,M., Watanabe,M., Hiraoka,S., Ishii,S., Kawai,Y., Saito,K., Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagahari,K., Masuho,Y. and Kanehori,K.		
	NEDO human cDNA sequencing project		
	Unpublished (2000)		
	2 (bases 1 to 2020)		
	Isogai,T. and Otsuki,T.		
	Direct Submission		
	Submitted (23-AUG-2000) to the DDBJ/EMBL/GenBank databases. Takao Isogai, Helix Research Institute, Genomics Laboratory. 1532-3 Yana Kisarazu, Chiba 292-0812, Japan (E-mail:genomics@hri.co.jp, Tel:81-438-52-3951, Fax:81-438-52-3952)		
	NEDO human cDNA sequencing project supported by Ministry of International Trade and Industry of Japan; CDNA full insert sequencing: Research Association for Biotechnology; CDNA library construction. 5'- & 3'-end one pass sequencing and clone selection Helix Research Institute (supported by Japan Key Technology Center etc.) and Department of Virology, Institute of Medical Science, University of Tokyo.		
FEATURES	Location/Qualifiers		
source	1..2020		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone="OVARC1000106"		
	/clone-lib="OVARC1"		
	/tissue_type="ovary, tumor tissue"		
	/note="cloning vector: PME18SFL3"		
	160..924		
	/note="unnamed protein product"		
	/codon_start=1		
	/protein_id="BAB14512.1"		
	/db_xref="GI:10435164"		
CDS			

[illegible]

CC * 31036 31111: gap of unknown length
CC * 31112 31878: contig of 767 bp in length
CC * 31879 31954: gap of unknown length
CC * 31955 32718: contig of 764 bp in length
CC * 32719 32794: gap of unknown length
CC * 32795 32861: contig of 167 bp in length
CC * 32962 33037: gap of unknown length
CC * 33038 33077: contig of 40 bp in length.
XX
FH Key Location/Qualifiers
FH source 1. 33077
FT /db_xref="taxon:9606"
FT /organism="Homo sapiens"
FT /clone="B118"
XX
SQ Sequence 33077 BP; 6694 A; 8214 C; 8318 G; 7030 T; 2821 other;

Query Match 6.7%; Score 191.4; DB 44; Length 33077;
Best Local Similarity 97.0%; Pred. No. 9.8e-29;
Matches 195; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1355 gggcagttatggcaggtgttatttaaacattctatgttaagtgtgttggc 1414
DB 31956 GCGCTGTTATGACACGCTTATTTAAACATTTCTATGTAATGTAATGTTTGGCT 32015
QY 1415 tctatgcacgttgagcagctggggggaagagaactgaccaggttaattgaatgaccc 1474
DB 32016 TTTACTGCTTTGTGAGCGCTGGGGGGAAGAGAACTGACCCAGTAATGTAATGAGGCC 32075
QY 1475 ttccctggaactaacagtccttgatgtgtgtgactaagtaagaatgataaacccatc 1534
DB 32076 TTCCCTGGAACTAACACGCTTGATGTTGTGTGACTAGTAAGATGTAATACCCCATC 32135
QY 1535 tgcctgggggtgtcactcaaca 1555
DB 32136 TGCTGGGGGTGTCACCTTCATA 32156

RESULT 9
166494/c 166494 7218 bp DNA PAT 28-DEC-1997
LOCUS
DEFINITION Sequence 14 from patent US 5670367.
ACCESSION 166494
VERSION 166494.1 GI:2724471
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 7218)
AUTHORS Donner, F., Scheiflinger, F. and Falkner, F. Gunter.
TITLE Recombinant fowlpox virus.
JOURNAL Patent: US 5670367-A 14 23-SEP-1997;
FEATURES
Location/Qualifiers
1. 7218
/organism="unknown"
BASE COUNT 1944 a 1491 c 1486 g 1929 t 368 others
ORIGIN

Query Match 3.2%; Score 92; DB 81; Length 7218;
Best Local Similarity 6.1%; Pred. No. 1.3e-08;
Matches 26; Conservative 254; Mismatches 144; Indels 0; Gaps 0;

QY 245 tcaaaagaggttgtgtcctgagcagacatagatgaacgagcgaagaagaggaagaag 304
DB 1459 TTAAGAGATGAGAAATTTGCTACRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1400
QY 305 aatggagaaggtcgaagaactgcaagatccagaagaatgtccagaagaggtttatgac 364
DB 1399 RRR 1340

QY 365 ctcatctctatataaaggctacaggaacagagagcagcagcagagtgacgag 424
DB 1339 RRR 1280
QY 425 aacagttcaattcaaaaacatgtaagagcgttgatgataagatgagaccattcctg 484
DB 1279 RRR 1220
QY 485 atgaggttctcgagcagcactaatagaaaacagcagaagaagaactgaaag 544
DB 1219 RRR 1160
QY 545 aactgaaggaatcacgaataacctaagaaggttgaatttctcaagaagaagaag 604
DB 1159 RRR 1100
QY 605 aagtggaaaagaactgtgactgtgaagcctatgaaccagaacaagttcccgagga 664
DB 1099 RRR 1040
QY 665 agct 668
DB 1039 ACCT 1036

RESULT 10
AL391825/c 201341 bp DNA HTG 06-SEP-2000
LOCUS Homo sapiens chromosome 1 clone RP11-456P18, *** SEQUENCING IN
DEFINITION PROGRESS ***, 29 unordered pieces.
ACCESSION AL391825
VERSION AL391825.3 GI:10040176
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 201341)
AUTHORS Plumb, B.
TITLE Direct Submision
JOURNAL Submitted (06-SEP-2000) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Sep 9, 2000 this sequence version replaced gi:994368.
COMMENT
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: BA456P18
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: Plasmid; 108752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 186960 bases at least Q40
Consensus quality: 192367 bases at least Q30
Consensus quality: 195566 bases at least Q20
Insert size: 198541; sum-of-contigs
Insert size: 216452; 5.7% error; agarose-ff
Quality coverage: 3.51x in Q20 bases; sum-of-contigs Quality
coverage: 3.36x in Q20 bases; agarose-ff

* NOTE: This is a 'working draft' sequence. It currently
* consists of 29 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 2205: contig of 2205 bp in length
* 2206 2305: gap of 100 bp

```
* 2306 6052: contig of 3747 bp in length
* 6053 6152: gap of 100 bp
* 6153 8931: contig of 2779 bp in length
* 8932 9031: gap of 100 bp
* 9032 15094: contig of 6063 bp in length
* 15095 15194: gap of 100 bp
* 15195 26633: contig of 11439 bp in length
* 26634 26733: gap of 100 bp
* 26734 32874: contig of 6141 bp in length
* 32875 32974: gap of 100 bp
* 32975 39594: contig of 6620 bp in length
* 39595 39694: gap of 100 bp
* 39695 42232: contig of 2538 bp in length
* 42233 42332: gap of 100 bp
* 42333 48607: contig of 6275 bp in length
* 48608 48707: gap of 100 bp
* 48708 52628: contig of 3921 bp in length
* 52629 52728: gap of 100 bp
* 52729 67221: contig of 14493 bp in length
* 67222 67321: gap of 100 bp
* 67322 73811: contig of 6490 bp in length
* 73812 73911: gap of 100 bp
* 73912 85339: contig of 11428 bp in length
* 85340 85439: gap of 100 bp
* 85440 95820: contig of 10381 bp in length
* 95821 95920: gap of 100 bp
* 95921 101396: contig of 5476 bp in length
* 101397 101496: gap of 100 bp
* 101497 108427: contig of 6931 bp in length
* 108428 108527: gap of 100 bp
* 108528 112215: contig of 3688 bp in length
* 112216 112315: gap of 100 bp
* 112316 114657: contig of 2342 bp in length
* 114658 114757: gap of 100 bp
* 114758 132207: contig of 17450 bp in length
* 132208 132307: gap of 100 bp
* 132308 148296: contig of 15989 bp in length
* 148297 148396: gap of 100 bp
* 148397 153491: contig of 5095 bp in length
* 153492 153591: gap of 100 bp
* 153592 157784: contig of 4193 bp in length
* 157785 157884: gap of 100 bp
* 157885 160643: contig of 2759 bp in length
* 160644 160743: gap of 100 bp
* 160744 165095: contig of 4352 bp in length
* 165096 165195: gap of 100 bp
* 165196 171176: contig of 5981 bp in length
* 171177 171276: gap of 100 bp
* 171277 179900: contig of 8624 bp in length
* 179901 180000: gap of 100 bp
* 180001 184362: contig of 4362 bp in length
* 184363 184462: gap of 100 bp
* 184463 191214: contig of 6752 bp in length
* 191215 191314: gap of 100 bp
* 191315 201341: contig of 10027 bp in length.

FEATURES
Source
1.201341
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/clone_lib="RP11-456P18"
/clone_1fb="RP11-11.2"
1.2205
/note="assembly_fragment:01082
clone_end:SP6
vector_side:left"
misc_feature
2306..6052
/note="assembly_fragment:00659
fragment_chain:1"
misc_feature
6153..8931
/note="assembly_fragment:00506
fragment_chain:1"
misc_feature
9032..15094
/note="assembly_fragment:00616
fragment_chain:1"
misc_feature
15195..26633
/note="assembly_fragment:01635
fragment_chain:1"
misc_feature
26734..32874
/note="assembly_fragment:01125
fragment_chain:1"
misc_feature
32975..39594
/note="assembly_fragment:00639
fragment_chain:1"
misc_feature
39695..42232
/note="assembly_fragment:00878
fragment_chain:2"
misc_feature
42333..48607
/note="assembly_fragment:01344
fragment_chain:2"
misc_feature
48708..52628
/note="assembly_fragment:00284
fragment_chain:2"
misc_feature
52729..67221
/note="assembly_fragment:00228
fragment_chain:2"
misc_feature
67322..73811
/note="assembly_fragment:00235
fragment_chain:3"
misc_feature
73912..85339
/note="assembly_fragment:00329
fragment_chain:3"
misc_feature
85440..95820
/note="assembly_fragment:00393
fragment_chain:3"
misc_feature
95921..101396
/note="assembly_fragment:00629
fragment_chain:4"
misc_feature
101497..108427
/note="assembly_fragment:01217
fragment_chain:4"
misc_feature
108528..112215
/note="assembly_fragment:00389
fragment_chain:4"
misc_feature
112316..114657
/note="assembly_fragment:01181
fragment_chain:5"
misc_feature
114758..132207
/note="assembly_fragment:01139
fragment_chain:5"
misc_feature
132308..148296
/note="assembly_fragment:00491
fragment_chain:5"
misc_feature
148397..153491
/note="assembly_fragment:00556
fragment_chain:6"
misc_feature
153592..157784
/note="assembly_fragment:00922
fragment_chain:6"
misc_feature
157885..160643
/note="assembly_fragment:00452"
fragment_chain:6"
misc_feature
160744..165095
/note="assembly_fragment:00582"
fragment_chain:6"
misc_feature
165196..171176
/note="assembly_fragment:01005"
fragment_chain:6"
misc_feature
171277..179900
/note="assembly_fragment:01295"
fragment_chain:6"
misc_feature
180001..184362
/note="assembly_fragment:01314"
fragment_chain:6"
misc_feature
184463..191214
/note="assembly_fragment:01341
fragment_chain:7"
misc_feature
191315..201341
/note="assembly_fragment:00355
fragment_chain:7"
misc_feature
clone_end:T7
```



```
*          91948 92047: gap of 100 bp
*          92048 114234: contig of 22187 bp in length
*          114235 114334: gap of 100 bp
*          114335 137242: contig of 22908 bp in length
*          137243 137342: gap of 100 bp
*          137343 162093: contig of 24751 bp in length.
FEATURES
source      1..162093
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="1"
            /map="1"
            /clone="RP11-195G14"
            /clone_lib="NPC1-11 Human Male BAC"
            1..2191
            /note="assembly_fragment"
            2292..3768
            /note="assembly_fragment"
            clone_end:SP6
            vector_side:left"
misc_feature 3869..6178
            /note="assembly_fragment"
            6279..11566
            /note="assembly_fragment"
            11667..16603
            /note="assembly_fragment"
            16704..21450
            /note="assembly_fragment"
            21551..28322
            /note="assembly_fragment"
            28423..35617
            /note="assembly_fragment"
            35718..47209
            /note="assembly_fragment"
            47310..57529
            /note="assembly_fragment"
            57630..73525
            /note="assembly_fragment"
            73626..91947
            /note="assembly_fragment"
            clone_end:T7
            vector_side:right"
misc_feature 92048..114234
            /note="assembly_fragment"
            114335..137242
            /note="assembly_fragment"
            137343..162093
            /note="assembly_fragment"
BASE COUNT 49114 a 31642 c 30768 g 49157 t 1412 others
ORIGIN
```

```
|||||
Db 100058 AAAAAAANCANGTTAA----- 10004
QY 498 acagcaggaactaataagaagaagaagaagaactgaagaactgaagaata 557
Db 100042 ----AGGCTTAACCTGGAAGAAAGCATCAAGGAGAAAGAGGAAATGAGGAAGT 99988
QY 558 cagaataaactcaagaagttggaattttcagaagaagaagaagaatggaagaag 617
Db 99987 TAGGAAGATGATCTTAACAACTGAGAAATATCT-TTGAACCAACAAAGGAATATGAGAGAA 99929
QY 618 actgactgaagcctaataagaagaagaagaacttcctccagcggaagctgtggcag 677
Db 99928 ACTGGCTGTAAATTTTAAACCAAGAACTATTTCTTTAGACAAA--ATTGGCATG 99871
QY 678 agctgtgaagcataagagctcagaagtggaacagtggtgaagaagctgaaccgagcc 737
Db 99870 AACAGTCAATATTTAAAGCCAGAAACACAGATGATTC-AAAAGCTGAATTTGTCC 99812
QY 738 tgaagcagatgaagaatcaagaagccctcatctcgaagctcctcgaa 787
Db 99811 TGAATCAAAATGTTGGAAATGAGATGTTTCATATTTGTTATTCAAA 99762

RESULT 12
LOCUS AC017643 177816 bp DNA HTG 10-DEC-1999
DEFINITION Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
pieces.
ACCESSION AC017643
AC017643.1 GI:6554354
KEYWORDS HTG: HTGS_PHASE2.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 177816)
AUTHORS Adams,M. and Venter,J.C.
TITLE Direct Submission
JOURNAL Submitted (10-DEC-1999) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA
COMMENT This sequence was identified as CDW:10211585 by the submitter.
For more information on this record e-mail to fly@celera.com.
* NOTE: This is a 'working draft' sequence.
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
FEATURES
source      1..177816
            /organism="Drosophila melanogaster"
            /db_xref="taxon:7227"
BASE COUNT 50527 a 38300 c 39085 g 49904 t
ORIGIN
```

Db 49774 GAACCCCTCAACAAATAGAGCAAGACATGAGTTTCAGAGAGCCCAAGCTG 49833
Oy 439 aaaaac 445
Db 49834 AGTAGA 49840

RESULT 13
AE003802
LOCUS
DEFINITION Drosophila melanogaster genomic scaffold 14200001386047 section 41
of 52, complete sequence.
ACCESSION AE003802 AE002787
VERSION AE003802.2 GI:10727480
KEYWORDS
SOURCE HTG.
ORGANISM fruit fly
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 262395)
AUTHORS Adams, M.D., Celniker, S.E., Holt, R.A., Evans, C.A., Gocayne, J.D.,
Amann-Edwards, P.G., Scherer, S.E., Li, P.W., Hoskins, R.A., Galle, R.F.,
George, R.A., Lewis, S.E., Richards, S., Ashburner, M., Henderson, S.N.,
Barton, G.C., Wortman, J.R., Yandell, M.D., Zhang, Q., Chen, L.X.,
Brudnon, R.C., Rogers, Y.H., Blazer, R.G., Champe, M., Pfeiffer, B.D.,
Man, K.H., Doyle, C., Baxter, E.G., Helt, G., Nelson, C.R., Gabor
Milos, G.L., Abrell, J.F., Agbayani, A., An, H.J.,
Andrews, P., Plambeck, C., Baldwin, D., Balley, R.M., Basu, A.,
Baxendale, J., Bayraktaroglu, L., Beasley, E.M., Beeson, K.Y.,
Benos, P.V., Berman, B.P., Bhandari, D., Bolshakov, S., Borkov, D.,
Bouchan, M.R., Bouck, J., Brokstein, P., Brotlier, P., Burris, K.C.,
Busam, D.A., Butler, H., Cadieu, E., Center, A., Chandra, I.,
Cherry, J.M., Cawley, S., Dahlke, C., Davenport, L.B., Davies, P., de
Pablos, B., Delcher, A., Deng, Z., Mays, A.D., Dew, I., Dietz, S.M.,
Dodson, K., Dou, L.E., Downes, M., Dugan-Kocha, S., Dunkov, B.C.,
Dunn, P., Durbin, K.J., Evangelista, C.C., Ferraz, C., Ferrier, S.,
Fleischmann, W., Foster, C., Gabriellian, A.E., Garg, N.S.,
Gelbart, W.M., Glasser, K., Glodek, A., Gong, F., Gorrell, J.H., Gu, Z.,
Guan, P., Harris, M., Harris, N.L., Harvey, D., Helman, T.J.,
Hernandez, J.R., Houck, J., Hostin, D., Houston, K.A., Howland, T.J.,
Wei, M.H., Idegawa, C., Jalali, M., Kalush, F., Karpen, G.H., Ke, Z.,
Kennisson, J.A., Ketchum, K.A., Kimmel, B.E., Kodira, C.D., Kraft, C.,
Kravitz, S., Kulp, D., Lai, Z., Lasko, P., Lei, Y., Levitsky, A.A.,
Li, J., Li, Z., Liang, Y., Lin, X., Liu, X., Matel, B., McIntosh, T.C.,
McLeod, M.P., McPherson, D., Merkulov, G., Milshina, N.V., Mobarry, C.,
Morris, J., Moshrefi, A., Mount, S.M., Moy, M., Murphy, B., Murphy, L.,
Muzny, D.M., Nelson, D.L., Nelson, D.R., Nelson, K.A., Nixon, K.,
Nusshorn, D.R., Paclet, J.M., Palazozo, M., Pittman, G.S., Pan, S.,
Pollard, D., Puri, V., Reese, M.G., Reinert, K., Remington, K.,
Saunders, R.D., Scheeler, F., Shen, H., Shue, B.C., Siden-Kiamos, I.,
Stapleton, M., Skupski, M.P., Smith, T., Spier, E., Spradling, A.C.,
Stapleton, M., Strong, R., Sun, E., Svirskas, R., Tector, C., Turner, R.,
Venter, E., Wang, A.H., Wang, X., Wang, Z.Y., Wassarman, D.A.,
Weissenbach, G.M., Weissenbach, J., Williams, S.M., Woodage, T.,
Worley, K.C., Wu, D., Yang, S., Yao, Q.A., Ye, J., Yeh, R.F.,
Zaveri, J.S., Zhan, M., Zhang, G., Zhao, Q., Zheng, L., Zheng, X.H.,
Zhong, F.N., Zhong, W., Zhou, X., Zhu, S., Zhu, X., Smith, H.O.,
Gibbs, R.A., Myers, E.W., Rubin, G.M. and Venter, J.C.
The genome sequence of *Drosophila melanogaster*
Science 287 (5461), 2185-2195 (2000)
20196006

TITLE 2 (bases 1 to 262395)
JOURNAL MEDLINE
REFERENCE
AUTHORS Adams, M.D., Celniker, S.E., Gibbs, R.A., Rubin, G.M. and Venter, C.J.
TITLE Direct Submission
JOURNAL Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA
COMMENT On Oct 9, 2000 this sequence version replaced gi:7302706.
FEATURES
SOURCE location/Qualifiers
1..262395
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"

mRNA

/chromosome="2R"
<1..27
/gene="CG4878"
/product="CT43034"
/db_xref="FLYBASE:FBan0004878"
/db_xref="FLYBASE:FBgn0034237"
<1..27
/gene="CG4878"
/product="CT15671"
/db_xref="FLYBASE:FBan0004878"
/db_xref="FLYBASE:FBgn0034237"
<1..>27
/gene="CG4878"
/map="54C10-54C12"
/db_xref="FLYBASE:FBan0004878"
/db_xref="FLYBASE:FBgn0034237"
join(<246..1745,1804..>2757)
/gene="CG18432"
/product="CT41966"
/db_xref="FLYBASE:FBan0018432"
/db_xref="FLYBASE:FBgn0034238"
/evidence=not_experimental
<246..>2757
/gene="CG18432"
/map="54C10-54C12"
/db_xref="FLYBASE:FBan0018432"
/db_xref="FLYBASE:FBgn0034238"
/evidence=not_experimental
join(246..1745,1804..2757)
/gene="CG18432"
/note="CG18432 gene product"
/codon_start=1
/db_xref="FLYBASE:FBan0018432"
/db_xref="FLYBASE:FBgn0034238"
/evidence=not_experimental
/protein_id="AA57841.1"
/db_xref="GI:7302762"
/translation="MEALSRKTPAKKRRKTHTRGVRKRRREOVOKLRATEGRISPE
HDMALSLFHFEGFEGEEOENDPAVEDCOEDSDSGELKAEPTDPSOTPIHDLH
DPLAWMOQKORHYVDESILLYRSNTLHQLHLSLSEKDSQITRLNVOOSKSK
SGKRGSLIQLQSGOKPASTFLMGKHLIRSVCHRRFOVFFPPMPRLKLPKSRICK
LCKQLNFSHPQLLEHRLAQQLLELDVYDSQOKRORIKRLKEELEIKRQVALKTL
ASAQDQTVAEKRLQKLTQLRQRYVESASAQRLQLRLKELKAKLKELRQORFOCT
RFQGLDYVHPPELAEASYCAWKSEFEDLAEVYEHLEFLYRLRLMTDOKSRFOCT
GHSKPPKPPQFDRIMASLRKREYDFDPEPEYVRLNADEXTARLSIPGGDLPK
ARNEFKIFPLDGGVGGOSRIYRELPDQISINECTIGVLRSLPENITVLYEKSTLT
PKSRRLAQISRPVLSKSDAVQKLSFQALSSQLAGDVYLYEYVSTGMSALHL
DDVORLPDALRQTLPLPQSLPALPQASKELVTPRPSGRIRKSQSLPIFDEQQLQFCG
LDVLDLRRRQLLHRSRQQRNLHTRQALFVPALEQETSEEPPLPDGNGTVQLPEPTV
WNPFDLHKHGRKFLQLLYEVIASQSAARAKALQTPLLLLAEDSPRSSGSGTWTALM
RPFCSVROGGRNSLPREPSCWGHQSAADLFKSPCVSLHYVRATGVPRRHILNLNE
RRSSAGDMSLVEPTOSE"

mRNA

join(<2926..3204,3268..3993,4054..4292,4416..>4506)
/gene="CG18631"
/product="CT41968"
/db_xref="FLYBASE:FBan0018631"
/db_xref="FLYBASE:FBgn0034239"
/evidence=not_experimental
<2926..>4506
/gene="CG18631"
/map="54C10-54C12"
/db_xref="FLYBASE:FBan0018631"
/db_xref="FLYBASE:FBgn0034239"
/evidence=not_experimental
join(2926..3204,3268..3993,4054..4292,4416..4506)
/gene="CG18631"
/note="CG18631 gene product"
/codon_start=1
/db_xref="FLYBASE:FBan0018631"
/db_xref="FLYBASE:FBgn0034239"
/evidence=not_experimental
/protein_id="AA57840.1"
/db_xref="GI:7302762"

CDS

1..262395
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"

2109 2188: gap of unknown length
2189 2885: contig of 697 bp in length
2189 2965: gap of unknown length
2189 4067: contig of 1102 bp in length
2189 4147: gap of unknown length
2189 4148 4828: contig of 681 bp in length
2189 4829 4908: gap of unknown length
2189 4909 5593: contig of 685 bp in length
2189 5594 5673: gap of unknown length
2189 5674 6235: contig of 562 bp in length
2189 6236 6315: gap of unknown length
2189 6316 7284: contig of 969 bp in length
2189 7285 7365 8031: contig of 666 bp in length
2189 8031 8111 8980: contig of 870 bp in length
2189 8981 9060: gap of unknown length
2189 10159: contig of 1099 bp in length
2189 10160 10239: gap of unknown length
2189 10240 11519: contig of 1280 bp in length
2189 11520 11599: gap of unknown length
2189 11600 12456: contig of 857 bp in length
2189 12457 12536: gap of unknown length
2189 12537 13277: contig of 741 bp in length
2189 13278 13357: gap of unknown length
2189 13358 14226: contig of 869 bp in length
2189 14227 14306: gap of unknown length
2189 14307 15161: contig of 855 bp in length
2189 15162 15241: gap of unknown length
2189 15242 16288: contig of 1047 bp in length
2189 16289 16368: gap of unknown length
2189 16369 17437: contig of 1069 bp in length
2189 17438 17517: gap of unknown length
2189 17518 18435: contig of 918 bp in length
2189 18436 18515: gap of unknown length
2189 18516 19746: contig of 1231 bp in length
2189 19747 19826: gap of unknown length
2189 19827 20681: contig of 855 bp in length
2189 20682 20761: gap of unknown length
2189 20762 21781: contig of 1020 bp in length
2189 21782 21861: gap of unknown length
2189 21862 22251: contig of 1390 bp in length
2189 22252 23331: gap of unknown length
2189 23332 24141: contig of 810 bp in length
2189 24142 24221: gap of unknown length
2189 24222 25604: contig of 1383 bp in length
2189 25605 25684: gap of unknown length
2189 25685 26897: contig of 1213 bp in length
2189 26898 26977: gap of unknown length
2189 26978 28192: contig of 1215 bp in length
2189 28193 28272: gap of unknown length
2189 28273 29096: contig of 824 bp in length
2189 29097 29176: gap of unknown length
2189 29177 29884: contig of 708 bp in length
2189 29885 29964: gap of unknown length
2189 29965 31202: contig of 1238 bp in length
2189 31203 31282: gap of unknown length
2189 31283 32902: contig of 1620 bp in length
2189 32903 32982: gap of unknown length
2189 32983 34609: contig of 1627 bp in length
2189 34610 34689: gap of unknown length
2189 34690 36022: contig of 1333 bp in length
2189 36023 36102: gap of unknown length
2189 36103 37659: contig of 1557 bp in length
2189 37660 37739: gap of unknown length
2189 37740 38822: contig of 1083 bp in length
2189 38823 38902: gap of unknown length
2189 38903 40562: contig of 1660 bp in length
2189 40563 40642: gap of unknown length
2189 40643 41706: contig of 1064 bp in length
2189 41707 41786: gap of unknown length
2189 41787 43869: contig of 2083 bp in length
2189 43870 43949: gap of unknown length
43950 46439: contig of 2490 bp in length
43950 46440 46519: gap of unknown length
43950 46520 49019: contig of 2500 bp in length
43950 49020 49099: gap of unknown length
43950 49100 51851: contig of 2752 bp in length
43950 51852 51931: gap of unknown length
43950 51932 53655: contig of 1724 bp in length
43950 53656 55864: contig of 2129 bp in length
43950 55865 55944: gap of unknown length
43950 55945 58995: contig of 3051 bp in length
43950 58996 59073: gap of unknown length
43950 59076 62709: contig of 3634 bp in length
43950 62710 62789: gap of unknown length
43950 62790 67945: contig of 5156 bp in length
43950 67946 68025: gap of unknown length
43950 68026 71131: contig of 3106 bp in length
43950 71132 71211: gap of unknown length
43950 71212 75476: contig of 4265 bp in length
43950 75477 75556: gap of unknown length
43950 75557 79689: contig of 4133 bp in length
43950 79690 79770 83653: contig of 3884 bp in length
43950 83654 83733: gap of unknown length
43950 83734 88122: contig of 4389 bp in length
43950 88123 88202: gap of unknown length
43950 88203 93246: contig of 5044 bp in length
43950 93247 93326: gap of unknown length
43950 93327 101880: contig of 8554 bp in length
43950 101881 101960: gap of unknown length
43950 101961 102656: contig of 696 bp in length
43950 102657 102736: gap of unknown length
43950 102737 103419: contig of 663 bp in length
43950 103420 103499: gap of unknown length
43950 103500 104124: contig of 625 bp in length
43950 104125 104204: gap of unknown length
43950 104205 104890: contig of 666 bp in length
43950 104891 104970: gap of unknown length
43950 104971 105607: contig of 637 bp in length
43950 105608 105687: gap of unknown length
43950 105688 106356: contig of 669 bp in length
43950 106357 106436: gap of unknown length
43950 106437 107123: contig of 687 bp in length
43950 107124 107204 107832: gap of unknown length
43950 107205 107913 108362: contig of 629 bp in length
43950 108363 108442: gap of unknown length
43950 108443 109244: contig of 802 bp in length
43950 109245 109325 110138: gap of unknown length
43950 109326 110218: contig of 814 bp in length
43950 110219 110239 110400: gap of unknown length
43950 110401 111120: contig of 822 bp in length
43950 111121 111814: contig of 694 bp in length
43950 111815 111894: gap of unknown length
43950 111895 112635: contig of 741 bp in length
43950 112636 112715: gap of unknown length
43950 112716 113419: contig of 704 bp in length
43950 113420 113499: gap of unknown length
43950 113500 114266: contig of 767 bp in length
43950 114267 114346: gap of unknown length
43950 114347 115075: contig of 729 bp in length
43950 115076 115155: gap of unknown length
43950 115156 115792: contig of 637 bp in length
43950 115793 115872: gap of unknown length
43950 115873 116391: contig of 519 bp in length
43950 116392 116471: gap of unknown length
43950 116472 117156: contig of 685 bp in length
43950 117157 117236: gap of unknown length
43950 117237 117918: contig of 682 bp in length
43950 117919 117998: gap of unknown length
43950 117999 118596: contig of 598 bp in length

Query Match	2.6%;	Score 74;	DB 29;	Length 129891;
Best Local Similarity	61.3%;	Pred. No. 7.3e-05;		
Matches 119; Conservative	0;	Mismatches 75;	Indels 0;	Gaps 0;

QY	252	gaattttctgcctgagcgagaaactagatgaaagcgcgcaaaagagcgaaagaatttgga	311
Db	13224	GGGCTTTTGTGACTGAAGCTGAGCGCGACCCGAGCAAAAGCAGACACTCAGAGGAAATTGGCA	13175
QY	312	gaaaagtcgaaaacctgtagaatccagaagaatgltccagaagaggtttagaacctcagtc	371
Db	13174	GGGTGTACGCCACGCGCGAGGATTCACCTGGAAGCAGACTGTGAGGAACCTTCAGATGGCGGCTC	13115
QY	372	tctatatagaagagctacaggaacagaaagacagaagcagcagaagtcagagaaacagtt	431
Db	13114	ACTTTATATACGCCCTCCAAACAAATTAAGACAGAAGGACATGAGATTCTGAGGAAGCCCA	13055
QY	432	caaatctcaaaaaca	445
Db	13054	CAAGCTGAGTAAAGA	13041

RESULT	15
HSMB02260	
LOCUS	
DEFINITION	HSMB02260 337 bp mRNA
	Homo sapiens mRNA; CDNA DKFZp434P1018 (from clone DKFZp434P1018), partial cds.
	18-FEB-2000

FEATURES

	Location/Qualifiers
1.	
2.	
3.	
4.	
5.	
6.	
7.	
8.	
9.	
10.	
11.	
12.	
13.	
14.	
15.	
16.	
17.	
18.	
19.	
20.	
21.	
22.	
23.	
24.	
25.	
26.	
27.	
28.	
29.	
30.	
31.	
32.	
33.	
34.	
35.	
36.	
37.	
38.	
39.	
40.	
41.	
42.	
43.	
44.	
45.	
46.	
47.	
48.	
49.	
50.	
51.	
52.	
53.	
54.	
55.	
56.	
57.	
58.	
59.	
60.	
61.	
62.	
63.	
64.	
65.	
66.	
67.	
68.	
69.	
70.	
71.	
72.	
73.	
74.	
75.	
76.	
77.	
78.	
79.	
80.	
81.	
82.	
83.	
84.	
85.	
86.	
87.	
88.	
89.	
90.	
91.	
92.	
93.	
94.	
95.	
96.	
97.	
98.	
99.	
100.	

BASE COUNT	1004 a	657 c	636 g	1030 t
ORIGIN				

Query Match	2.6%	Score 73.4;	DB 78;	Length 3327;
Best Local Similarity	70.5%;	Pred. No. 7.3e-05;		
Matches 98;	Conservative 0;	Mismatches 41;	Indels 0;	Gaps 0;

QY	2712	cccttcgctggaagaagctgaacgaggggttcgaaggttc	aaacacccatcaat	2771
Db	3161	CTTTAGTGTTGATTAATAAAGTCACGTGCTGCACACCA	AAAAAAAAAAAAAAAA	3220
QY	2772	gaagcgaatgaacctgcagagaaacaaataaaaaaataa	aaaaaataaaaaaa	2833
Db	3221	AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA	AAAAAAAAAAAAAAAA	3280
QY	2832	aaaaaaaaaaaaaaaaaa	2850	
Db	3281	AAAAAAAAAAAAAAAAAAAA	3299	

```
Search completed: March 1, 2001, 02:41:45
Job time: 25111 sec
```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 28, 2001, 22:49:15 : Search time 234.97 Seconds
(without alignments)
4556.491 Million cell updates/sec

Title: US-09-602-597-1
Perfect score: 2850
Sequence: 1 aagccctgaaggtcaaaag.....aaaaaaaaaaaaaaaaaaaaa 2850

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapept 1.0

Searched: 480022 seqs, 187831343 residues

Total number of hits satisfying chosen parameters: 960044

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

N.Geneseq-36.*
1: /SID56/gcgdata/geneseq/geneseq/NA1980.DAT.*
2: /SID56/gcgdata/geneseq/geneseq/NA1981.DAT.*
3: /SID56/gcgdata/geneseq/geneseq/NA1982.DAT.*
4: /SID56/gcgdata/geneseq/geneseq/NA1983.DAT.*
5: /SID56/gcgdata/geneseq/geneseq/NA1984.DAT.*
6: /SID56/gcgdata/geneseq/geneseq/NA1985.DAT.*
7: /SID56/gcgdata/geneseq/geneseq/NA1986.DAT.*
8: /SID56/gcgdata/geneseq/geneseq/NA1987.DAT.*
9: /SID56/gcgdata/geneseq/geneseq/NA1988.DAT.*
10: /SID56/gcgdata/geneseq/geneseq/NA1989.DAT.*
11: /SID56/gcgdata/geneseq/geneseq/NA1990.DAT.*
12: /SID56/gcgdata/geneseq/geneseq/NA1991.DAT.*
13: /SID56/gcgdata/geneseq/geneseq/NA1992.DAT.*
14: /SID56/gcgdata/geneseq/geneseq/NA1993.DAT.*
15: /SID56/gcgdata/geneseq/geneseq/NA1994.DAT.*
16: /SID56/gcgdata/geneseq/geneseq/NA1995.DAT.*
17: /SID56/gcgdata/geneseq/geneseq/NA1996.DAT.*
18: /SID56/gcgdata/geneseq/geneseq/NA1997.DAT.*
19: /SID56/gcgdata/geneseq/geneseq/NA1998.DAT.*
20: /SID56/gcgdata/geneseq/geneseq/NA1999.DAT.*
21: /SID56/gcgdata/geneseq/geneseq/NA2000.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1817.2	63.8	2011	233669	Human breast tumor
2	114	4.0	117	T22042	Human gene signal
3	70.8	2.5	882	252528	Human secreted pro
4	68.6	2.4	2773	21265009	Membrane-bound pro
5	68.4	2.4	3508	21265009	Human secreted pro
6	68.2	2.4	1034	21265009	Human secreted pro
7	68.2	2.4	2846	21265009	Membrane-bound pro
8	68	2.4	1153	202655748	Human secreted pro
9	68	2.4	1086	16087587	DNA encoding Leuco
10	67.8	2.4	4186	20241320	Human normal ovary
11	67.8	2.4	2123	18259701	PTH-like peptide D
12	67.8	2.4	12827	19209036	Equine arteritis v

13	67.8	2.4	15528	19209039	Equine arteritis v
14	67.6	2.4	1307	17263176	CDNA from clone dx
15	67.2	2.4	1641	17215758	Human interleukin
16	67	2.4	2082	19215257	Mouse neuronal PAS
17	66.8	2.3	2262	21252560	Human secreted pro
18	66.6	2.3	441	20209474	p135-NT3 construct
19	66.6	2.3	1024	21297189	Human prostate can
20	66.6	2.3	1052	21263358	Human secreted pro
21	66.6	2.3	4120	20209473	Human RNA helicase
22	66.4	2.3	2300	20200686	Human GPC3 DNA. H
23	66.4	2.3	2764	19263190	CDNA from clone dh
24	66.4	2.3	3637	20260802	Human secreted pro
25	66.4	2.3	4061	19215927	Clone cr797.3 codi
26	66	2.3	1104	20215927	Human gene express
27	66	2.3	1248	2126436	Human secreted pro
28	66	2.3	1299	21216666	Human secreted pro
29	66	2.3	2400	11200505	Placenta-specific
30	66	2.3	3347	9280639	Sequence of CDNA e
31	65.8	2.3	1210	21250064	Impatiens balsamia
32	65.8	2.3	1361	21263399	Human secreted pro
33	65.8	2.3	1673	20277320	Human secreted pro
34	65.8	2.3	4237	19261487	Human secreted pro
35	65.8	2.3	4650	20233987	Human secreted pro
36	65.8	2.3	6010	122011998	Human PRO708 nucle
37	65.8	2.3	6020	112006648	Human plasminogen
38	65.8	2.3	7753	152056826	Plasminogen gene f
39	65.6	2.3	719	19264501	Notwalk virus stra
40	65.6	2.3	772	1926363	Human secreted pro
41	65.6	2.3	790	19263117	Coding sequence fo
42	65.6	2.3	790	21239903	Inserted DNA fragm
43	65.6	2.3	986	20251735	DNA fragment from
44	65.6	2.3	2685	21243781	DNA encoding a hum
45	65.4	2.3	1954	21243441	Human fetal brain
					CDNA encoding huma

ALIGNMENTS

RESULT 1	
ID 233669	233669 standard; CDNA; 2011 BP.
AC 233669;	
DT 08-DEC-1999	(first entry)
DE Human breast tumour-associated EST 59.	
KW Expressed sequence tag; EST; human; breast; cancer; gene therapy; treatment; tumour; cytostatic; medicament; ss.	
OS Homo sapiens.	
XX DE19813839-A1.	
PN 23-SEP-1999.	
PD 20-MAR-1998;	98DE-1013839.
PF 20-MAR-1998;	98DE-1013839.
PR 20-MAR-1998;	98DE-1013839.
XX (META-) METAGEN GES GENOMFORSCHUNG MBH.	
PA Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E, Rosenthal A;	
PI WPI, 1999-528981/45.	
XX P-PSDB; Y48610.	
DR Human nucleic acid sequences and protein products from tumor breast	
XX tissue, useful for breast cancer therapy -	
PT Claim 1a; 136-137; 188pp; German.	
XX	

T22042		standard; cDNA to mRNA; 117 BP.
ID	T22042	
XX		
AC	T22042;	
XX		
DT	07-AUG-1996	(first entry)
XX		
DE	Human gene signature HUMGS03585.	
KM	Gene signature; messenger RNA; mRNA; relative abundance; frequency;	
KW	human; cloning; mapping; non-biased library; diagnosis; detection;	
XX	cell typing; abnormal cell function; ss.	
OS	Homo sapiens.	
XX		
PN	WO9514772-A1.	
XX		
PD	01-JUN-1995.	
XX		
PF	11-NOV-1994; 94WO-JP01916.	
XX		
PR	12-NOV-1993; 93JP-0355504.	
XX		
PA	(MATS/) MATSUBARA K.	
XX	(OKUBO/) OKUBO K.	
PI	Matsubara K, Okubo K;	
DR	WPI: 1995-206931/27.	
XX		
PT	Identifying gene signatures in 3'-directed human cDNA library - e.g.	
PT	for diagnosis of abnormal cell function, by preparing cDNA that	
PT	reflects relative abundance of corresp. mRNA in specific human	
XX	tissues	
PS	Claim 1; Page 1030; 2245pp; Japanese.	
XX		
CC	A single-stranded DNA (or its complementary strand or the corresp.	
CC	double-stranded DNA) which comprises one of the 7837 "GS" sequences	
'CC	given in T19001-T26837 and which is able to hybridise to part of	
CC	human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)	
CC	sequences were obtained from 3'-directed cDNA libraries prepared	
CC	from various human tissues; synthesis of cDNA was initiated from the	
CC	3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-	
CC	untranslated sequence is unique to a particular mRNA species, almost	
CC	all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library	
CC	is constructed so as to reflect accurately the relative abundance of	
CC	different mRNAs in the particular tissue from which it was derived.	
CC	The appearance frequency of a given GS in a cDNA library can be	
CC	determined (esp. using primers and probes derived from the GS	
CC	sequences) as a means of diagnosing abnormal cell function or for	
XX	recognising different cell types.	
SEQ	Sequence 117 BP; 38 A; 30 C; 15 G; 32 T; 2 other;	
Query Match	4.0%; Score 114; DB 16; Length 117;	
Best Local Similarity	98.3%; Pred. No. 1.2e-14;	
Matches 114; Conservative	0; Mismatches 2; Indels 0; Gaps 0;	
OY	1018 gatcgacgcttcattcgtgcgtactatagcatgatctccctcaaaaaaaactccttgcct	1077
DB	1 gatcgacgcttcattcgtgcgtactatagcatgatctccctcaaaaaaaactccttncct	60
OY	1078 gatacttgtagtaacaatgacatttttaaccaatccaatcaaaaatgtgccaga	1133
DB	61 gatacttgtagtaacaatgacatttttaaccaatccaatcaaaaatgtgccaga	116
RESULT	3	
ID	Z52528 standard; cDNA; 882 BP.	
XX		

AC Z52528.
DT 29-FEB-2000 (first entry)
DE Human secreted protein clone yB187_1 nucleotide sequence SEQ ID NO:107.
XX
XX
KW Human; secreted protein; immunostimulatory; haemostatic; cytokine;
KM proliferative; differentiative; chemotactic; chemokinetic; vaccine;
NM thrombolytic; antiinflammatory; cytotoxic; immunosuppressive;
gene therapy; ss.
OS Homo sapiens.
PN WO955642-AZ.
PD 18-NOV-1999.
PF 14-MAY-1999; 99MO-US10843.
PI 14-MAY-1998; 98US-0085472.
PR 17-AUG-1998; 98US-0096824.
PR 11-SEP-1998; 98US-0099843.
PR 11-SEP-1998; 98US-0099950.
PR 15-SEP-1998; 98US-0100424.
PR 29-SEP-1998; 98US-0102329.
PR 09-OCT-1998; 98US-0103615.
PR 11-DEC-1998; 98US-0111799.
PR 14-DEC-1998; 98US-0112159.
PR 31-DEC-1998; 98US-0114415.
PR 10-FEB-1999; 99US-0248059.
PR 06-APR-1999; 99US-0287150.
PR 13-MAY-1999; 99US-0311021.

(GENEY) GENETICS INST INC.
PA
XX
XX
PL Wong GG, Clark HF, Fechtel K, Agostino MJ;
DR WPT: 2000-053095/04.
P-PSDB: Y734A3.
XX Novel polynucleotides and proteins having biological activities which
PT make them suitable for treating, preventing or ameliorating medical
PP conditions in humans or animals -
PX
XX Claim 116; Page 654; 730pp; English.
PS
XX The present invention describes human secreted proteins encoded by
CC polynucleotides obtained from adult testes, foetal brain, adult brain,
CC brain (foetal and adult), foetal kidney, adult spleen, and adult thymus
CC cDNA libraries. The polynucleotides and proteins are predicted to have
CC biological activities which would make them suitable for treating,
CC preventing or ameliorating medical conditions in humans and animals.
CC Suggested activities include nutritional activity, cytokine and cell
CC proliferation/differentiation activity, immune stimulating (e.g. as
CC vaccines) or suppressing activity, hematopoiesis regulating activity,
CC tissue growth activity, activin/inhibin activity, chemotactic/
CC chemokinetec activity, haemostatic and thrombolytic activity, receptor/
CC ligand activity, anti-inflammatory activity, cadherin/tumour invasion
CC suppressor activity, and tumour inhibition activity. The polynucleotides
CC are also stated to be useful for gene therapy. Therapeutic compositions
CC are also presently valuable for veterinary applications. Z52475 to Z52581
CC encode human secreted proteins, and Y73390 to Y73500 represent human
CC secreted proteins, given in the present invention.
XX
SO Sequence 882 BP; 253 A; 225 C; 209 G; 195 T; 0 other;

Query Match 2.5%; Score 70.8; DB 21; Length 882;
Best Local Similarity 57.1%; Pred.No.1.3e-05;
Matches 129; Conservative 0; Mismatches 97; Indels 0; Gaps 0;

| 2625 cacatgttacaaggcagccacgaacataagaatcatcatacttatgtagcaaacgc 2684

PR 17-AUG-1998; 98US-0096867.
 PR 17-AUG-1998; 98US-0096891.
 PR 17-AUG-1998; 98US-0096894.
 PR 17-AUG-1998; 98US-0096895.
 PR 17-AUG-1998; 98US-0096897.
 PR 18-AUG-1998; 98US-0096949.
 PR 18-AUG-1998; 98US-0096950.
 PR 18-AUG-1998; 98US-0096959.
 PR 18-AUG-1998; 98US-0096960.
 PR 18-AUG-1998; 98US-0097022.
 PR 19-AUG-1998; 98US-0097141.
 PR 20-AUG-1998; 98US-0097218.
 PR 20-AUG-1998; 98US-0097611.
 PR 26-AUG-1998; 98US-0097951.
 PR 26-AUG-1998; 98US-0097952.
 PR 26-AUG-1998; 98US-0097954.
 PR 26-AUG-1998; 98US-0097955.
 PR 26-AUG-1998; 98US-0097971.
 PR 26-AUG-1998; 98US-0097971.
 PR 26-AUG-1998; 98US-0097974.
 PR 26-AUG-1998; 98US-0097978.
 PR 26-AUG-1998; 98US-0097979.
 PR 26-AUG-1998; 98US-0097986.
 PR 26-AUG-1998; 98US-0098014.
 PR 31-AUG-1998; 98US-0098525.
 PR 16-SEP-1998; 98US-0100634.
 PR 12-JAN-1999; 99US-0115565.

XX (GENT) GENENTECH INC.

XX Baker K, Chen J, Goddard A, Gurney AL, Smith V, Watanabe CK;
 PI Wood WI, Yuan J;
 XX

DR MPI: 2000-072883/06.
 DR P-PSDB; Y66674.

XX Membrane-bound proteins and related nucleotide sequences

XX Claim 2; Fig 112; 822pp; English.

CC The invention provides membrane-bound PRO polypeptides and
 CC polynucleotides encoding them. The PRO sequences of the invention were
 CC identified based on extracellular domain homology screening. The PRO
 CC sequences have homology with proteins including LDL receptors, TIE
 CC ligands and various enzymes. The membrane-bound proteins and receptor
 CC molecules are useful as pharmaceutical and diagnostic agents. Receptor
 CC immunoadhesins, for instance, can be used as therapeutic agents to block
 CC receptor-ligand interactions. The membrane-bound proteins can also be
 CC employed for screening of potential peptide or small molecule inhibitors
 CC of the relevant receptor/ligand interaction. The PRO encoding sequences
 CC are useful as hybridization probes, in chromosome and gene mapping and in
 CC the generation of antisense RNA and DNA. PRO nucleic acid sequences
 CC will also be useful for the preparation of PRO polypeptides, especially
 CC by recombinant techniques.

XX Sequence 2773 BP; 832 A; 652 C; 650 G; 639 T; 0 other;

Query Match 2.4%; Score 68.6; DB 21; Length 2773;
 Best Local Similarity 70.2%; Pred. No. 5.2e-05;
 Matches 92; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

QY 2720 tctgggaagaagtcgacgaggttgcaagggttaaaacatcatcaatgaagcta 2779
 Db 2624 tctgtaattaaatcgtttagagatgtaaaaaataaaaaataaaaaataaaaaa 2683
 QY 2780 ataacctgtcagagaaacaaaaataaaaaataaaaaataaaaaataaaaaa 2839
 Db 2684 aaaaaataaaaaataaaaaataaaaaataaaaaataaaaaataaaaaataaaaaa 2743
 QY 2840 aaaaaaaaaa 2850
 Db 2744 aaaaaaaaaa 2754

RESULT 5

Al6619 ID Al6619 standard; CDNA, 3508 BP.

XX Al6619;

DT 16-JUN-2000 (first entry)

DE Human secreted protein clone l0311_8 nucleotide sequence SEQ ID NO:3.

XX Human; secreted protein; immunostimulant; immunosuppressant; virucide;
 KW antibacterial; antifungal; cytostatic; antiarthritic; antirheumatic; dermatologic;
 KW antidiabetic; antiasymptomatic; antitumor; antiparasitic; antiprotocole;
 KW antihypertensive; immunodeficiency; severe combined immunodeficiency; SCID;
 KW infection; HIV; hepatitis; malaria; autoimmune disorder; systemic lupus;
 KW connective tissue disease; multiple sclerosis; erythematosis;
 KW rheumatoid arthritis; autoimmune pulmonary inflammation; asthma;
 KW Guillain-Barre syndrome; autoimmune thyroiditis; myasthenia gravis;
 KW insulin dependent diabetes mellitus; graft-versus-host-disease;
 KW autoimmune inflammatory eye disease; allergy; ss.

OS Homo sapiens.

PN WO200009552-A1.

PD 24-FEB-2000.

PF 13-AUG-1999; 99WO-US18298.

XX 14-AUG-1998; 98US-0096622.
 PR 17-AUG-1998; 98US-0096815.
 PR 04-SEP-1998; 98US-0099229.
 PR 23-OCT-1998; 98US-0105368.
 PR 08-JAN-1999; 99US-0115234.
 PR 12-FEB-1999; 99US-0119931.
 PR 18-FEB-1999; 99US-0120575.
 PR 30-APR-1999; 99US-0132020.
 PR 11-AUG-1999; 99US-0096622.

XX (GENT) GENENTECH INC.

PI Jacobs K, McCoy JM, Lavallie ER, Collins-Racie LA, Evans C;
 PI Merberg D, Tracy M, Agostino MJ, Steinhilber RJ, Spaulding V;
 PI Wong GG, Clark HF, Fechtel K;

XX MPI: 2000-205979/18.
 DR P-PSDB; Y94899.

XX New polynucleotides encoding secreted proteins, which may have e.g.
 PT nutritional, chemokine, immune stimulating or suppressing,
 PT hematopoiesis regulating, tissue growth, activin/inhibin
 PT antiinflammatory or tumor inhibition activity
 XX Claim 12; Page 467-468; 641pp; English.

CC Al6618 to Al6619 encode the human secreted proteins given in Y94898
 CC to Y94980, isolated from human adult brain, adult thyroid, adult retina,
 CC foetal carcinoma, adult blood, adult neural, foetal kidney, adult
 CC placenta, adult testis, whole embryo, adult cartilage, kidney, foetal
 CC brain, adult thymus, foetal placenta, adult uterus, adult tumour, and
 CC adult bladder, CDNA libraries. The polynucleotides and proteins are
 CC predicted to have biological activities which would make them suitable
 CC for treating, preventing or ameliorating medical conditions in humans
 CC and animals. The polynucleotides can be used as markers for tissues in
 CC which the protein is preferentially expressed, as molecular weight
 CC markers on Southern gels, and as chromosome markers or tags to identify
 CC chromosomes or to map gene positions. The proteins can be used in the
 CC treatment of immune deficiencies and disorders, such as severe combined
 CC immunodeficiency (SCID), as well as viral, bacterial, fungal and other
 CC infections. These infections include human immunodeficiency virus (HIV),
 CC hepatitis, herpesviruses, mycobacteria, Leishmania spp., malaria and
 CC candidiasis. The proteins can be used to treat autoimmune disorders such


```

PR 04-JUN-1998; 98US-0088033.
PR 04-JUN-1998; 98US-0088326.
PR 05-JUN-1998; 98US-0088167.
PR 05-JUN-1998; 98US-0088202.
PR 05-JUN-1998; 98US-0088212.
PR 05-JUN-1998; 98US-0088217.
PR 09-JUN-1998; 98US-0088655.
PR 10-JUN-1998; 98US-0088722.
PR 10-JUN-1998; 98US-0088730.
PR 10-JUN-1998; 98US-0088734.
PR 10-JUN-1998; 98US-0088738.
PR 10-JUN-1998; 98US-0088740.
PR 10-JUN-1998; 98US-0088741.
PR 10-JUN-1998; 98US-0088742.
PR 10-JUN-1998; 98US-0088810.
PR 10-JUN-1998; 98US-0088811.
PR 10-JUN-1998; 98US-0088824.
PR 10-JUN-1998; 98US-0088825.
PR 10-JUN-1998; 98US-0088826.
PR 11-JUN-1998; 98US-0088858.
PR 11-JUN-1998; 98US-0088861.
PR 11-JUN-1998; 98US-0088863.
PR 11-JUN-1998; 98US-0088876.
PR 12-JUN-1998; 98US-0089030.
PR 12-JUN-1998; 98US-0089105.
PR 16-JUN-1998; 98US-0089440.
PR 16-JUN-1998; 98US-0089512.
PR 16-JUN-1998; 98US-0089514.
PR 17-JUN-1998; 98US-0089532.
PR 17-JUN-1998; 98US-0089538.
PR 17-JUN-1998; 98US-0089598.
PR 17-JUN-1998; 98US-0089599.
PR 17-JUN-1998; 98US-0089600.
PR 17-JUN-1998; 98US-0089601.
PR 18-JUN-1998; 98US-0089603.
PR 18-JUN-1998; 98US-0089607.
PR 18-JUN-1998; 98US-0089908.
PR 19-JUN-1998; 98US-0089947.
PR 19-JUN-1998; 98US-0089948.
PR 19-JUN-1998; 98US-0089952.
PR 22-JUN-1998; 98US-0090246.
PR 22-JUN-1998; 98US-0090252.
PR 22-JUN-1998; 98US-0090254.
PR 23-JUN-1998; 98US-0090349.
PR 23-JUN-1998; 98US-0090355.
PR 24-JUN-1998; 98US-0090429.
PR 24-JUN-1998; 98US-0090431.
PR 24-JUN-1998; 98US-0090435.
PR 24-JUN-1998; 98US-0090444.
PR 24-JUN-1998; 98US-0090445.
PR 24-JUN-1998; 98US-0090461.
PR 24-JUN-1998; 98US-0090472.
PR 24-JUN-1998; 98US-0090538.
PR 24-JUN-1998; 98US-0090538.
PR 24-JUN-1998; 98US-0090540.
PR 24-JUN-1998; 98US-0090557.
PR 25-JUN-1998; 98US-0090676.
PR 25-JUN-1998; 98US-0090678.
PR 25-JUN-1998; 98US-0090688.
PR 25-JUN-1998; 98US-0090690.
PR 25-JUN-1998; 98US-0090691.
PR 25-JUN-1998; 98US-0090694.
PR 25-JUN-1998; 98US-0090695.
PR 25-JUN-1998; 98US-0090696.
PR 26-JUN-1998; 98US-0090862.
PR 26-JUN-1998; 98US-0090863.
PR 01-JUL-1998; 98US-0091358.
PR 01-JUL-1998; 98US-0091360.
PR 01-JUL-1998; 98US-0091544.
PR 02-JUL-1998; 98US-0091478.
PR 02-JUL-1998; 98US-0091486.
PR 02-JUL-1998; 98US-0091519.
PR 02-JUL-1998; 98US-0091626.

PR 02-JUL-1998; 98US-0091628.
PR 02-JUL-1998; 98US-0091633.
PR 02-JUL-1998; 98US-0091646.
PR 02-JUL-1998; 98US-0091673.
PR 07-JUL-1998; 98US-0091978.
PR 07-JUL-1998; 98US-0091982.
PR 09-JUL-1998; 98US-0092182.
PR 10-JUL-1998; 98US-0092472.
PR 20-JUL-1998; 98US-0093339.
PR 30-JUL-1998; 98US-0094651.
PR 04-AUG-1998; 98US-0095282.
PR 04-AUG-1998; 98US-0095285.
PR 04-AUG-1998; 98US-0095301.
PR 04-AUG-1998; 98US-0095302.
PR 04-AUG-1998; 98US-0095318.
PR 04-AUG-1998; 98US-0095321.
PR 04-AUG-1998; 98US-0095325.
PR 10-AUG-1998; 98US-0095325.
PR 10-AUG-1998; 98US-0095929.
PR 10-AUG-1998; 98US-0096012.
PR 11-AUG-1998; 98US-0096143.
PR 11-AUG-1998; 98US-0096173.
PR 11-AUG-1998; 98US-0096791.
PR 12-AUG-1998; 98US-0096329.
PR 17-AUG-1998; 98US-0096757.
PR 17-AUG-1998; 98US-0096766.
PR 17-AUG-1998; 98US-0096768.
PR 17-AUG-1998; 98US-0096773.
PR 17-AUG-1998; 98US-0096791.
PR 17-AUG-1998; 98US-0096867.
PR 17-AUG-1998; 98US-0096891.
PR 17-AUG-1998; 98US-0096894.
PR 17-AUG-1998; 98US-0096895.
PR 17-AUG-1998; 98US-0096897.
PR 18-AUG-1998; 98US-0096949.
PR 18-AUG-1998; 98US-0096950.
PR 18-AUG-1998; 98US-0096959.
PR 18-AUG-1998; 98US-0096960.
PR 18-AUG-1998; 98US-0097022.
PR 19-AUG-1998; 98US-0097141.
PR 20-AUG-1998; 98US-0097218.
PR 24-AUG-1998; 98US-0097661.
PR 26-AUG-1998; 98US-0097951.
PR 26-AUG-1998; 98US-0097952.
PR 26-AUG-1998; 98US-0097954.
PR 26-AUG-1998; 98US-0097955.
PR 26-AUG-1998; 98US-0097971.
PR 26-AUG-1998; 98US-0097974.
PR 26-AUG-1998; 98US-0097978.
PR 26-AUG-1998; 98US-0097979.
PR 26-AUG-1998; 98US-0097986.
PR 26-AUG-1998; 98US-0098014.
PR 31-AUG-1998; 98US-0098525.
PR 16-SEP-1998; 98US-0100634.
PR 12-JAN-1999; 99US-0115565.

XX
XX
PA (GETH ) GENENTECH INC.
XX
XX Baker K, Chen J, Goddard A, Gurney AL, Smith V, Watanabe CK;
PI Wood WI, Yuan J;
XX
XX WPI: 2000-072883/06.
DR P-PSDB: Y66695.
DR
XX
XX Membrane-bound proteins and related nucleotide sequences
XX
XX
XX Claim 2: Fig 158; 822pp; English.
PS
CC The invention provides membrane-bound PRO polypeptides and
CC polynucleotides encoding them. The PRO sequences of the invention were
CC identified based on extracellular domain homology screening. The PRO
CC sequences have homology with proteins including LDL receptors, TIE
CC ligands and various enzymes. The membrane-bound proteins and receptor
CC molecules are useful as pharmaceutical and diagnostic agents. Receptor

```


DR P-PSDB; R70491.
 XX
 PT Immune inducing polypeptide against Leucocytozoan protozoa - useful
 PT in production of vaccines for treatment of leucocytozoanosis in
 PT fowl.
 PS Claim 1; Page 12-14; 20pp; Japanese.
 XX
 CC Q87587-89 encode polypeptides having a whole or partial epitope of a
 CC structural protein of Leucocytozoan protozoa (see R70491-93). The
 CC polypeptides and DNA encoding them are useful in the production of
 CC vaccines for the treatment of leucocytozoanosis of fowl.
 XX
 SQ Sequence 1686 BP; 915 A; 129 C; 382 G; 260 T; 0 other;

Query Match 2.4%; Score 68; DB 16; Length 1686;
 Best Local Similarity 49.8%; Pred. No. 5.9e-05;
 Matches 202; Conservative 0; Mismatches 200; Indels 4; Gaps 1;

QY 264 tgaagcgaactagatgaacgcgcgcaaaagagcaagaagatgagagaagttcgaaa 323
 DB 222 tgaagaaagaagaagaagaagaagaagaagaagaagaagaagaagaagaaga 281
 QY 324 accggaagatccagaagaatgtccagagaggtttatgaccctcgatctctatatgaag 383
 DB 282 agaaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaaga 341
 QY 384 gctacgaagacagaagagcaggaagcagaggtacgaggaagaagatcgaatcaaaa 443
 DB 342 tgaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaaga 401
 QY 444 catggtgaaggtctgaatgaagatgagaccacttccttgatgaggtttctcgacgca 503
 DB 402 aaatgaagatgaacaataatgaatgaacaataatgaagaagaagaagaagaaga 461
 QY 504 ggaactatgaaagaagaagaagaagaagaagaagaagaagaagaagaagaagaaga 559
 DB 462 agaaacaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaaga 521
 QY 560 gaataacctcaagaaggttgaatttctcaagaagaagaagaagaagaagaagaaga 619
 DB 522 agaaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaaga 581
 QY 620 tgactgtgaagcctatagaaccagaagaagttctccagggcgaa 665
 DB 582 tgaagatgaagaaaaaaagaagaagaagaagaagaagaagaagaatgaa 627

RESULT 10
 Z41320
 ID Z41320 standard; CDNA; 4186 BP.
 XX
 AC Z41320;
 XX
 DT 18-JAN-2000 (first entry)
 XX
 DE Human normal ovarian tissue derived cDNA 99.
 XX
 KW Human; ovary; screening; ovarian cancer; treatment; ss.
 XX
 OS Homo sapiens.
 XX
 PN DE19816395-A1.
 XX
 PD 07-OCT-1999.
 XX
 PE 03-APR-1998; 98DE-1016395.
 XX
 PR 03-APR-1998; 98DE-1016395.
 XX
 PA (META-) METAGEN GDS GENOMFORSCHUNG MBH.
 XX

PI Rosenthal A, Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E;
 XX
 DR WPI; 1999-552352/47.
 XX
 PT Nucleic acid sequences potentially useful in diagnosis or therapy of
 PT ovarian cancer
 XX
 PS Claim 3; Page 209-210; 274ppp; German.
 XX
 CC This invention describes novel nucleic acid sequences that are highly
 CC expressed in normal ovary tissue. Artificial chromosomes and cosmid
 CC clones containing the sequences can be used as gene transfer vehicles.
 CC The sequences can be used to produce DNA fragments containing full-length
 CC genes. Host cells transformed with the sequences can be used to produce
 CC polypeptides or polypeptide fragments, which can be used to screen phage
 CC displays for polypeptides that bind to them, or as tools for identifying
 CC agents active against ovarian cancer, or to prepare medicaments for
 CC treating ovarian cancer. The cDNA sequences can be used to obtain genomic
 CC genes, their promoters, enhancers, silencers, exon structures, intron
 CC structures and their splice variants. Z41222-Z41324 represent cDNA
 CC sequences derived from normal human ovarian tissue and which encode the
 CC protein fragments represented in Y59724-Y59837.
 XX
 SQ Sequence 4186 BP; 938 A; 1199 C; 1234 G; 815 T; 0 other;

Query Match 2.4%; Score 68; DB 20; Length 4186;
 Best Local Similarity 76.9%; Pred. No. 7.8e-05;
 Matches 83; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

QY 2743 ttgcaggttttaaaatcattacatgaagaagctatgaactgtcagagaaacaaa 2802
 DB 2590 ttgaagtcgttaaaaaaaatgaagaagaagaagaagaagaagaagaagaaga 2649
 QY 2803 aaaaaaaatgaagaagaagaagaagaagaagaagaagaagaagaagaagaaga 2850
 DB 2650 aaaaaaaatgaagaagaagaagaagaagaagaagaagaagaagaagaagaaga 2697

RESULT 11
 T59701
 ID T59701 standard; DNA; 2123 BP.
 XX
 AC T59701;
 XX
 DT 07-JUL-1997 (first entry)
 XX
 DE PTH-like peptide DNA clone HHM-4.
 XX
 KW PTH-like peptide; parathyroid hormone; parathormone; agonist;
 KW adenylate cyclase; humoral hypercalcaemia of malignancy; HHM;
 KW tumour marker; cancer; diagnosis; ss.
 XX
 OS Homo sapiens.
 XX
 FH Key
 FT 5'UTR
 FT 1..938 Location/Qualifiers
 FT CDS
 FT 939..1547
 FT /tag= a
 FT /tag= b
 FT 939..1046
 FT sig_peptide
 FT /tag= c
 FT 1047..1544
 FT /tag= d
 FT 1545..2123
 FT /tag= e
 FT 2006..2011
 FT /tag= f
 FT 1062..1099
 FT /tag= g
 FT /note= "nucleotide region corresponding to probe
 FT used to screen SKRC-1 cDNA library"
 FT


```
XX 14-JUL-1997; 97WO-NL00408.
PF 12-JUL-1996; 96NL-0003579.
PR (UYLE-) RIJKSUNIV LEIDEN.
XX Bredenbeek PJ, Den Boon JA, Snijder EJ, Spaan WMJ;
PI Van Dinten LC, Wassenaar ALM;
XX WPI; 1998-110596/10.
DR
XX Recombinant DNA vector that produces infectious RNA of equine
PT arteritis virus - and derived RNA, host cells and virus particles,
PT useful therapeutically
XX
XX Claim 4; Pages 21-24; 36pp; English.
XX
XX The sequence is that of a recombinant DNA vector, pEAV030,
CC comprising a DNA sequence complementary to part of
CC the genome of equine arteritis virus (EAV) which has
CC adjacent to its 3'-end a non-EAV specific promoter for a non-EAV
CC specific RNA polymerase. It can be used to make viral particles
CC and for therapeutic use. The vector can be stably incorporated
CC into host cells and produces infectious RNA.
XX
SQ Sequence 15528 BP; 3492 A; 3941 C; 3972 G; 4123 T; 0 other;

Query Match 2.4%; Score 67.8; DB 19; Length 15528;
Best Local Similarity 62.9%; Pred. No. 0.00013;
Matches 105; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

QY 2684 cttgaagagccattactgtatctcattgaccttgatgtaagaagatgacagaaggt 2743
DB 12641 cgtgaagacggtatctctctgtgctgcatgtgtaagtagtattagccaccag 12700
QY 2744 tgcaggggtttaaaacatcattcaatgaagctataaactgtcagagacaacaaa 2803
DB 12701 aaccacaaataaaaaataaaaaataaaaaataaaaaataaaaaataaaaaa 12760
QY 2804 aaaaaataaaaaataaaaaataaaaaataaaaaataaaaaataaaaaa 2850
DB 12761 aaaaaataaaaaataaaaaataaaaaataaaaaataaaaaataaaaaa 12807

RESULT 14
V63176
ID V63176 standard; CDNA: 1307 BP.
XX
AC V63176;
DT 12-JAN-1999 (first entry)
DE CDNA from clone dx279_1 which encodes a secreted protein.
XX
KW Secreted protein; immune stimulating; suppressing;
KW haematopoiesis regulating activity; tissue growth activity; activin;
KW inhibin activity; chemotactic; chemokinetic activity; haemostatic;
KW thrombolytic activity; anti-inflammatory activity; cadherin;
KW tumour invasion suppressor activity; tumour inhibition activity; ds.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX CDS 449..1009
XX FT /*tag= a
XX PN WO9840486-A2.
XX PD 17-SEP-1998.
XX PF 13-MAR-1998; 98WO-US04977.
```

```
XX 29-OCT-1997; 97US-0960022.
PR 14-MAR-1997; 97US-0815047.
XX (GEMV ) GENETICS INST INC.
XX
XX Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D;
PI Racle LA, Spaulding V, Treacy M;
XX WPI; 1998-520812/44.
DR P-PSDB; W80402.
XX
XX New isolated human poly(nucleotide)s and secreted proteins -
PT obtained from e.g. human foetal kidney, placenta, foetal brain,
PT adult testes, adult brain or adult uterus CDNA libraries
XX
XX Claim 35; Pages 84-85; 110pp; English.
XX
XX The present sequence encodes a secreted protein. The nucleic acid
CC sequence is isolated from a human adult testes CDNA library using
CC probe V63186. The polypeptide may have biological activities such as
CC e.g. nutritional activity, immune stimulating or suppressing activity,
CC haematopoiesis regulating activity, tissue growth activity,
CC activin/inhibin activity, chemotactic/chemokinetic activity,
CC and thrombolytic activity, receptor/ligand activity, anti-inflammatory
CC activity, cadherin/tumour invasion suppressor activity, tumour inhibition
CC activity or other activities.
XX
SQ Sequence 1307 BP; 295 A; 442 C; 370 G; 200 T; 0 other;

Query Match 2.4%; Score 67.6; DB 19; Length 1307;
Best Local Similarity 80.6%; Pred. No. 6.5e-05;
Matches 79; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

QY 2752 ttaaaacatccattactgtaagcttaataaacctgtcagagacacaaaaa 2811
DB 1210 ttaaatcatcgcgttgcaaaaaataaaaaataaaaaataaaaaa 1269
QY 2812 aaaaaataaaaaataaaaaataaaaaataaaaaataaaaaa 2849
DB 1270 aaaaaataaaaaataaaaaataaaaaataaaaaataaaaaa 1307

RESULT 15
T15758
ID T15758 standard; CDNA: 1641 BP.
XX
AC T15758;
DT 24-JUL-1997 (revised)
DT 25-JUN-1996 (first entry)
DE Human interleukin-15 receptor CDNA from clone p1.
XX
KW interleukin-15 receptor; IL-15R; murine; human; graft rejection;
KW treatment; prevention; ss.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX CDS 3..842
XX FT /*tag= a
XX FT /product= interleukin-15_receptor
XX PN WO9530695-A.
XX PD 16-NOV-1995.
XX PF 04-MAY-1995; 95WO-US05585.
XX PF 06-SEP-1994; 94US-0300903.
```

PR 06-MAY-1994; 94US-0236919.

XX (IMMUNEX) IMMUNEX CORP.

XX PI Anderson DM, Giri JG;

XX DR WPI; 1996-077255/08.

XX P-PSDB; R90843.

PT DNA encoding an interleukin-15 receptor - used to ameliorate undesired
PT effects of interleukin-15, e.g. to prevent or treat graft rejection

XX Example 3; Page 35-37; 53pp; English.

XX A lambda-gt10 library from human peripheral blood lymphocytes, was
CC screened for the presence of a full-length clone encoding human
CC interleukin-15 receptor (IL-15R) using a random prime labeled human
CC IL-15R cDNA probe consisting of the entire W5 cDNA clone (799999)
CC without the polyA tail. The resulting cDNA insert from this clone
CC 15758 encodes R90843, and exhibited an in-frame insertion of 153 bp
CC at the mature N-terminus (amino acids 24 through 74), an in-frame
CC deletion of 99 bp downstream of the mature N-terminus that deleted
CC nucleotides 236 to 334 of T15995, and also contained additional 5'
CC sequence as compared to clone W5 (amino acids 1 through 10), but still
CC did not contain an initiator methionine. The IL-15R DNA is a multi-exon
CC gene. IL-15R variants can be attributed to different mRNA splicing events
CC following transcription or from proteolytic cleavage of the IL-15R
CC protein, where the IL-15R binding property is retained. The IL-15R
CC protein is useful in treatment of undesired effects of interleukin-15,
CC e.g. to prevent or treat graft rejection.
CC (Revised entry submitted to correct sequence analysis breakdown.)
XX
SQ Sequence 1641 BP; 478 A; 463 C; 385 G; 315 T; 0 other;

Query Match 2.4%; Score 67.2; DB 17; Length 1641;

Best Local Similarity 81.2%; Pred. No. 8.5e-05;

Matches 78; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

OY 2755 aaacatcattacatgaagctataaacctgcagagaaacaaaaaaaaaaaaa 2814

Db 1535 aaacatcatttcaataaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 1594

OY 2815 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 2850

Db 1595 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 1630

Search completed: March 1, 2001, 02:11:11
Job time: 12116 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 28, 2001, 18:20:14 ; Search time 3991.74 Seconds
(without alignments)
5003.167 Million cell updates/sec

Title: US-09-602-597-1
Perfect score: 2850
Sequence: 1 aagccctgaaggtcaaaa.....aaaaaaaaaaaaaaaaaa 2850

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 7991742 seqs, 3503743858 residues

Total number of hits satisfying chosen parameters: 15983484

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est25:*
26: gb_est26:*
27: gb_est27:*
28: gb_est28:*
29: gb_est29:*
30: gb_est30:*
31: gb_est31:*
32: gb_est32:*
33: gb_est33:*
34: gb_est34:*
35: gb_est35:*
36: gb_est36:*
37: gb_est37:*
38: gb_est38:*
39: gb_est39:*
40: gb_est40:*
41: em_est1a:*
42: em_est1b:*
43: em_est1c:*

44: em_est1d:*
45: em_est1e:*
46: em_est1f:*
47: em_est1g:*
48: em_est1h:*
49: em_est1i:*
50: em_est1j:*
51: em_est1k:*
52: em_est1l:*
53: em_est1m:*
54: em_est1n:*
55: em_est1o:*
56: em_est1p:*
57: em_est1q:*
58: em_est1r:*
59: em_est1s:*
60: em_est1t:*
61: em_est1u:*
62: em_est1v:*
63: em_est1w:*
64: em_est1x:*
65: em_est1y:*
66: em_est1z:*
67: em_est2a:*
68: em_est2b:*
69: em_est2c:*
70: em_est2d:*
71: em_est2e:*
72: em_est2f:*
73: em_est2g:*
74: em_est2h:*
75: em_est2i:*
76: em_est2j:*
77: em_est2k:*
78: em_est2l:*
79: em_est2m:*
80: em_est2n:*
81: em_est2o:*
82: em_est2p:*
83: em_est2q:*
84: em_est2r:*
85: em_est2s:*
86: em_est2t:*
87: em_est2u:*
88: em_est2v:*
89: em_est2w:*
90: em_est2x:*
91: em_est2y:*
92: em_est2z:*
93: em_est3a:*
94: em_est3b:*
95: em_est3c:*
96: em_est3d:*
97: em_est3e:*
98: em_est3f:*
99: em_est3g:*
100: em_est3h:*
101: em_est3i:*
102: em_est3j:*
103: em_est3k:*
104: em_est3l:*
105: em_est3m:*
106: em_est3n:*
107: em_est3o:*
108: em_est3p:*
109: em_est3q:*
110: em_est3r:*
111: em_est3s:*
112: em_est3t:*
113: em_est3u:*
114: em_est3v:*
115: em_est3w:*
116: em_est3x:*

117: em_estpl6:*

118: em_estpl7:*

119: em_estpl8:*

120: em_estpl9:*

121: em_estpl10:*

122: em_estpl11:*

123: em_estpl12:*

124: em_estpl13:*

125: em_estpl14:*

126: em_estpl15:*

127: em_estpl16:*

128: em_estpl17:*

129: em_estpl18:*

130: em_estpl19:*

131: em_estpl20:*

132: em_estpl21:*

133: em_estpl22:*

134: em_estpl23:*

135: em_estpl24:*

136: em_estpl25:*

137: em_estpl26:*

138: em_estpl27:*

139: em_estpl28:*

140: em_estpl29:*

141: em_estpl30:*

142: em_estpl31:*

143: em_estpl32:*

144: em_estpl33:*

145: em_estpl34:*

146: em_estpl35:*

147: em_estpl36:*

148: em_estpl37:*

149: em_estpl38:*

150: em_estpl39:*

151: em_estpl40:*

152: em_estpl41:*

153: em_estpl42:*

154: em_estpl43:*

155: em_estpl44:*

156: em_estpl45:*

157: em_estpl46:*

158: em_estpl47:*

159: em_estpl48:*

160: em_estpl49:*

161: em_estpl50:*

162: em_estpl51:*

163: em_estpl52:*

164: em_estpl53:*

165: em_estpl54:*

166: em_estpl55:*

167: em_estpl56:*

168: em_estpl57:*

169: em_estpl58:*

170: em_estpl59:*

171: em_estpl60:*

172: em_estpl61:*

173: em_estpl62:*

174: em_estpl63:*

175: em_estpl64:*

176: em_estpl65:*

177: em_estpl66:*

178: em_estpl67:*

179: em_estpl68:*

180: em_estpl69:*

181: em_estpl70:*

182: em_estpl71:*

183: em_estpl72:*

184: em_estpl73:*

185: em_estpl74:*

186: em_estpl75:*

187: em_estpl76:*

188: em_estpl77:*

189: em_estpl78:*

190: gb_gss25:*

191: gb_gss26:*

192: gb_gss27:*

193: gb_gss28:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	699.8	24.6	704	23	A1655644	A1655644 ts99g02.x
C 2	690.4	24.2	712	158	A0470436	A0470436 CTBT-ET-
C 3	606	21.3	632	135	BE727658	BE727658 601564388
C 4	596.2	20.9	614	95	AM821142	AM821142 PM2-ST030
C 5	594.2	20.9	614	95	BE387718	BE387718 601287377
C 6	594.2	20.9	614	95	BE732904	BE732904 601567847
C 7	588.2	20.6	699	136	BE869679	BE869679 601445783
C 8	583.4	20.5	664	107	BE392702	BE392702 601307895
C 9	581.8	20.4	631	105	BE180639	BE180639 RC3-HT062
C 10	575.2	20.2	895	135	BE736929	BE736929 601306864
C 11	564.8	19.8	568	89	AM327476	AM327476 dq03g09.x
C 12	552.4	19.4	554	92	AM592821	AM592821 h606c01.x
C 13	534.4	18.8	568	17	A1205174	A1205174 ap13a11.x
C 14	533.2	18.7	577	15	A1031969	A1031969 oW55d01.x
C 15	532.4	18.7	623	38	AV653573	AV653573 AV653573
C 16	519	18.2	519	136	BE858277	BE858277 7918g01.x
C 17	501.8	17.6	514	40	AM119015	AM119015 xq99b10.x
C 18	499	17.5	510	12	AA854150	AA854150 aj71f04.s
C 19	499	17.5	529	17	A1188440	A1188440 qd14b03.x
C 20	492.4	17.3	495	38	AM004593	AM004593 w890a04.x
C 21	489	17.2	489	26	A1863261	A1863261 tz45c12.x
C 22	487.8	17.1	491	9	AA594110	AA594110 tm81e09.s
C 23	482	16.9	482	19	A1346704	A1346704 qp47b12.x
C 24	481.6	16.9	627	95	AM821178	AM821178 PM2-ST030
C 25	476	16.7	476	25	A1819378	A1819378 w61d09.x
C 26	475	16.7	479	16	A1090352	A1090352 oy18b10.s
C 27	474.4	16.6	476	16	A1146389	A1146389 q805c07.x
C 28	469	16.5	481	105	BE233505	BE233505 601108191
C 29	467.8	16.4	471	25	A1799081	A1799081 w68d07.x
C 30	464.4	16.3	469	89	AM361848	AM361848 PM3-CY026
C 31	464	16.3	465	25	A1819590	A1819590 wk03a12.x
C 32	460.4	16.2	462	13	AA873225	AA873225 oH70g08.s
C 33	456	16.0	457	40	AM117393	AM117393 x800f07.x
C 34	452.8	15.9	499	39	AM084108	AM084108 xc37c08.x
C 35	450	15.8	475	89	AM327477	AM327477 dq03g09.y
C 36	449	15.8	449	21	A1479864	A1479864 tm70c12.x
C 37	448.8	15.7	472	137	BE3939496	BE3939496 OVO-UM009
C 38	448	15.7	448	134	BE047617	BE047617 tz39e03.y
C 39	448	15.7	471	142	H99248	H99248 yx21g06.s1
C 40	447.4	15.7	449	22	A1613225	A1613225 cy39g01.x
C 41	443	15.5	443	17	A1192097	A1192097 q606c09.x
C 42	442	15.5	444	9	AA598697	AA598697 ae49b05.s
C 43	441	15.5	453	16	A1148841	A1148841 qc65d12.x
C 44	439	15.4	439	18	A1278270	A1278270 q178b06.x
C 45	437	15.3	437	19	A1359389	A1359389 q728b06.x

ALIGNMENTS

RESULT 1

LOCUS A1655644

DEFINITION ts99g02.x1 NC1 CGAP GC6 Homo sapiens cDNA clone IMAGE:2239442 3'

ACCESSION A1655644

VERSION A1655644.1

KEYWORDS EST.

SOURCE human.

704 bp mRNA

GI:4739623

14-DEC-1999

ORGANISM Homo sapiens
 Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:
 Mammalia: Eutheria: Primates: Catarrhini: Homnidae: Homo.
 REFERENCE 1 (bases 1 to 704)
 AUTHORS NCI-CCGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1350
 Email: Robert.Strausberg@nih.gov
 Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
 R. Emmert-Buck, M.D., Ph.D.
 CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
 Bonaldo, Ph.D.
 CDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CCGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 www.bio.llnl.gov/dbp/image/image.html
 Inset Length: 844 Std Error: 0.00
 Seq primer: -400P from Gibco
 High quality sequence stop: 467.
 Location/Qualifiers
 1..704
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_image="2239442"
 /clone_lib="NCI-CCGAP-CC6"
 /tissue_type="pooled germ cell tumors"
 /lab_host="DH10B"
 /note="Vector: pT73D-Pac (Pharmacia) with a modified
 polylinker; Plasmid DNA from the normalized library
 NCI-CCGAP-CC6 was prepared, and ss circles were made in
 vitro. Following HAP purification, this DNA was used as
 tracer in a subtractive hybridization reaction. The driver
 was PCR-amplified cDNAs from a pool of 5,000 clones made
 from the same library (clones 1257096-1258651,
 1469064-1470983, and 1475592-1476743). Subtraction by
 Bento Soares and M. Fatima Bonaldo."
 BASE COUNT 139 a 156 c 186 g 222 t 1 others
 ORIGIN
 Query Match 24.6%; Score 699.8; DB 23; Length 704;
 Best Local Similarity 99.6%; Pred. No. 4e-141;
 Matches 701; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Db 344 TCCTGATGTCGCCCTTCATCCACTGCCCTGCTGCTCATATGTAACGCACTCTCCCA 285
 QY 853 ggcctgggtgctactcttgaggagcagcactccgagtcagtcagcagcagcagc 912
 Db 284 GGCTTGATGCTCACTCTGGAGCAGCAGCTCCAGTCAGTCAGCAGCAGCAGC 225
 QY 913 ataatgcaacccggaagatgtctctcatcttcgaaacacacttcctcgagcc 972
 Db 224 ATCAATGCCACCGAAGATGTCCTTCATCTCCGAAACCAACCTTCGAGCC 165
 QY 973 cctagttctcgcctccctacagagagtcctcccaaggtaagtcgagccgttc 1032
 Db 164 CCTAGTTCTCCGCTCCCTACACAGGAGCTCCTCCCAAGGTAGTCGAGCCGTTCA 103
 QY 1033 gctgcctatagcattatgtccctcaaaaaaactccttgctgcatctgtgtaca 1092
 Db 104 GCTGCTATAGGATATGTCCTCAAAAAAACTCTTGCTGCTGATCCTGTACAA 45
 QY 1093 catgacatttttaacatccatccataaaatgtgcagatcc 1136
 Db 44 CATGACATTTTAAACCAATCAATCAAAAAATGTGCCAGATCC 1

RESULT 2
 AQ470436/c 712 bp DNA GSS 23-APR-1999
 LOCUS CTRBI-EI-2584F21.TF CTRBI-EI Homo sapiens genomic clone 2584F21,
 DEFINITION DNA sequence.
 ACCESSION AQ470436
 VERSION AQ470436.1 GI:4654090
 KEYWORDS GSS.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:
 Mammalia: Eutheria: Primates: Catarrhini: Homnidae: Homo.
 REFERENCE 1 (bases 1 to 712)
 AUTHORS Zhao,S., Adams,M.D., Nieman,W., Malek,J., Shizuya,H., Simon,A. and
 Venter,J.C.
 TITLE Use of BAC End Sequences from Caltech Libraries for Sequence-Ready
 Map Building
 JOURNAL Unpublished (1997)
 COMMENT Other_GSSs: CTRBI-EI-2584F21.TF
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850
 Tel: 301 838 0208
 Fax: 301 838 0208
 Email: hbeet@igf.org
 Clones are available from Research Genetics (info@resgen.com). BAC
 end search page:
 http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
 Seg primer: M13-21
 Class: BAC ends.

FEATURES
 source Location/Qualifiers
 1..712
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_image="2584F21"
 /clone_lib="CTRBI-EI"
 /sex="male"
 /cell_type="sperm"
 /note="Vector: pBelOBAC11; Site_1: EcoRI; Site_2: EcoRI;
 Caltech Human BAC Library D"

BASE COUNT 210 a 157 c 171 g 174 t
 ORIGIN
 Query Match 24.2%; Score 690.4; DB 158; Length 712;
 Best Local Similarity 99.0%; Pred. No. 4.3e-139;
 Matches 705; Conservative 0; Mismatches 6; Indels 1; Gaps 1;

```

Oy 1747 caaacgttagccttcttactaatagccagacatcagaactgcaggttagtataatg 1806
      |||
Db 712 CAAAGTAGAGCCTGTTTACTAATAGCCAGACATCAGACTTCAGTAGATGTTAAG 653
      |||
Oy 1807 agatgaactattcttgcgagcctccggaaacctataatattgtaaatgagtgagacact 1866
      |||
Db 652 AAGATACCTATTCTTGGCAGCCTCGGAATCTAATATTGTAATGAGGAGACACACT 593
      |||
Oy 1867 gatatcttgaccattctatagagccc-tctctgttaatgatcatatattacttgctt 1925
      |||
Db 532 GCATTTTGGACCATTTCTATTGAGGCCCTTCTCTGTTTAATCATATTACTTGTGCTT 533
      |||
Oy 1926 ttaactgtggaactcatttcttaactaaagtgctgccttagtacttcttctgctcct 1985
      |||
Db 533 TTAAGTGTGGAATCTATTCTTAACCTAAAGTGTCGCCCTAGTACTTCTTGTGCGCT 473
      |||
Oy 1986 ctgcgtccttcttcttcccaacagcaactctgagccatgagcagcagcaaaactaga 2045
      |||
Db 472 CTGCTGCTCTTCTTCTTCCAAACAGCAACTCTGAGGCCATGAGCAGCAAAACTAGA 413
      |||
Oy 2046 ggtactgtccacccctgcctcatalaaggaacgggctcacccttgatcttgagagag 2105
      |||
Db 412 GGTACTGTCCACCTCGCTCTCATTAAGGAACGGGCTCATCTTGGATTCTGGAGAGAG 353
      |||
Oy 2106 ggaagaggagatggtgtggaagcctcgcagagagagatagatagacttgcacaacat 2165
      |||
Db 352 GGAGAGGGAGATGAGTGAGAGCCTCGAGAGACAGATAGACATGAGCTTTGACAACAT 293
      |||
Oy 2166 ctgtgagctctctgctttagaataagcatgtaacattcttcatcccttattctc 2225
      |||
Db 232 CTGTAGGCTCTCTCTCTTGAATAGCATGATCATCTTCTTATCCATCTCCCTTATTC 233
      |||
Oy 2226 tacacaaattggttacttcttctgtgtgagactgagtgagacacacacaatgtgt 2285
      |||
Db 232 TACATCAATTTGTTTACTTCTTGTGGTGTGAGACTGAGTGAACACACAAAATGTGT 173
      |||
Oy 2286 tgacactgtatgcccgcagcagcagcagctactgacttgaacatgggcagagagccc 2345
      |||
Db 172 TGACACTGTGATGCGCGAGCGAGCAGAGCTACTGACTTTGAACATGGGCAAGAGGCC 113
      |||
Oy 2346 ctgagctcattccacccactcttcttccctccaagtacagtgacactctgtgctccat 2405
      |||
Db 112 CTGGATATCATCCACCCACCTCTTTCCTCCCTCCAGTACAGTAAACACTCTGTGCCAT 53
      |||
Oy 2406 tggcagatgagcactccctgcagccataactgactgtgtaattcttc 2457
      |||
Db 52 TGGCAGATGGGACCTTCCTGCACCCATACATGATGCTTTGTGATTTGCCCC 1
      |||

RESULT 3
BE727698 632 bp mRNA EST 15-SEP-2000
LOCUS 601564388F1 NIH_MGC_20 Homo sapiens cDNA clone IMAGE:3834093 5',
DEFINITION mRNA sequence.
ACCESSION BE727698
VERSION BE727698.1 GI:10141690
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eulheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 632)
AUTHORS NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel.: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: ATCC/DC/DTP
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
```

```

FEATURES
    source
        1..632
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="IMAGE:3834093"
            /clone_1db="NIH_MGC_20"
            /tissue_type="melanotic melanoma"
            /lab_host="DH10B (phage-resistant)"
            /note="Organ: skin; Vector: pORF7; Site: 1: XhoI; Site: 2:
            EcoRI; cDNA made by oligo-dT priming. Directionally
            cloned into EcoRI/XhoI sites using the following 5'
            adaptor: GGCACGAG(G). Size-selected >500bp for average
            insert size 1.8kb. Library constructed by Ling Hong in
            the laboratory of Gerald M. Rubin (University of
            California, Berkeley) using ZAP-cDNA synthesis kit
            (Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT 210 a 168 c 140 g 114 t
ORIGIN

Query Match 21.3%; Score 606; DB 135; Length 632;
Best Local Similarity 100.0%; Pred. No. 7.6e-121;
Matches 606; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 475 aacttccttgatgaggttcttcgacagcagacactatagaagaacagcagagaaga 534
      |||
Db 1 AACTCTCTTGATGAGAGGTTCTTCGACAGCAGAACTAATAGAAACGACGAGAGAA 60
      |||
Oy 535 gaactgaagaactgaagaatatacagaataaccctcaagaaggttgaaattctcaagag 594
      |||
Db 61 GAACGGAAGAAGACTAAGGAATACGAATAAACCTCAAGAAGTTGGAATTTCTCAAGAG 120
      |||
Oy 595 aacaaagaagatgtaagaagaactgactgtgaagcctatagaacaacaaagaagtc 654
      |||
Db 121 AACCAAGAAGAGTGAAGAAAGAACTGACTGTGAAGCTTATAGAAACCAAGCAAGTTTC 180
      |||
Oy 655 tcccaaggagagctgttgccaggaagctgtgaagcctaagagctcgaagaaggcagagc 714
      |||
Db 181 TCCCAGGGAGAGCTTTTGCGAGAGCTGTGAAGCTAAGAGCTGAGAGAGTGCAACAGT 240
      |||
Oy 715 gtgaagaagactgaacccggaacctgagccagaatgaagaagaacagagccctcatctgc 774
      |||
Db 241 GTGAAGAAGACTGAACCGGAGCCTGAGCCAGATGACAAAGAAATCAAGACCCCTCATCTGC 300
      |||
Oy 775 aagtcctcgaagaacacctccctgagtgagccctccatccactgcctctgtgcagta 834
      |||
Db 301 AAGTCTCTCGAAGAACCTCTCTGAGTGAGGCCCTTCATCCACTGCCCTCTGCTGCAGTA 360
      |||
Oy 835 tttatcgcgacatccctccagagcctgtgtgctcactcttgaggaagcagactcgaagtcagc 894
      |||
Db 361 TGTATCGGATCTCTCCAGGCTGTGGGTCTACTCTTGGAAGAGCAGCATCCGATGCCAGC 420
      |||
Oy 895 tcaagaacaggaagcaccatcaatgccaacggaaagatgtctctcattctccgaacc 954
      |||
Db 421 TCAGACAGAGGAAGGACCATCAATGACCAACGGAAGATTTGCTCTCATGCTTCGGAAC 480
      |||
Oy 955 aacacctcctcgaagcctccctagttctcgcgtccctcacaaggaagctcctcccaagg 1014
      |||
Db 481 AACACCTTCTCTGAGGCCCTCTAGTTTCTCGTCCCTCAACAGGAGCTCCCTCCCAAGG 540
      |||
Oy 1015 gtagatggagcgttcatgctgacctataggaatattgcccctcaaaaaaaactcttct 1074
      |||
Db 541 GTAGATCGGAGCCGTTCATGCTGCTATAGGCAATTATGTCTCCCAAAAAAACTCTTTTG 600
      |||
Oy 1075 cctgca 1080
      |||
Db 601 CCTGCA 606
      |||

RESULT 4
```

AM821142
 LOCUS AM821142 614 bp mRNA EST 17-MAY-2000
 DEFINITION PM2-ST0303-030200-005-d12 ST0303 Homo sapiens cDNA, mRNA sequence.
 ACCESSION AM821142
 VERSION AM821142.1 GI:7914136
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 REFERENCE 1 (bases 1 to 614)
 Dias Neto,E., da Silva,Correa,R., Verjovski-Almeida,S., Brijones,M.R.,
 Nagai,M.A., da Silva,M., Jr., Zago,M.A., Bordin,S., Costa,F.F.,
 Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,
 Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
 M.J., Soares,F., Brentani,R.K., Reis,L.F., de Souza,S.J. and
 Simpson,A.J.
 Shotgun sequencing of the human transcriptome with ORF expressed
 sequence tags
 Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
 JOURNAL MEDLINE 20202663
 COMMENT Contact: Simpson A.J.G.
 Laboratory of Cancer Genetics
 Ludwig Institute for Cancer Research
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
 Brazil
 Tel: +55-11-2704922
 Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br
 This sequence was derived from the FAPESP/LICR Human Cancer Genome
 Project. This entry can be seen in the following URL
 (<http://www.ludwig.org.br/scripts/gethtml2.pl?l=6&t=PM2-ST0303-030>)
 200-005-d12&t3=2000-02-03&t4=1)
 Seq primer: puc 18 forward
 High quality sequence stop: 582.
 Location/Qualifiers
 1. 614
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="ST0303"
 /dev_stage="Adult"
 /note="Organ: stomach; Vector: puc18; Site_1: SmaI;
 Site_2: SmaI; A mini-library was made by cloning products
 derived from ORESTES PCR (U.S. Letters Patent application
 No. 196,716 - Ludwig Institute for Cancer Research)
 profiles into the puc 18 vector. Reverse transcription of
 tissue mRNA and cDNA amplification were performed under
 low stringency conditions."
 BASE COUNT 166 a 185 c 135 g 128 t
 ORIGIN

Query Match 20.9%; Score 596.2; DB 95; Length 614;
 Best Local Similarity 99.5%; Pred. No. 1e-118;
 Matches 598; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

|||||
 Db 254 CTGCTGAGATATGATGGGATCTCTCCAGGCGCTGGCTTCTATCTGGAGAGGCACT 313
 Qy 884 ccgagttccagctcagacagcgaaggaacacatcaatgaccacggaaagatgtctctcca 943
 Db 314 CCGAGTCAGCTCAGACAGGAGGACCAATCAATGACCCGGAAGATGTCTCTCCCA 373
 Qy 944 ttctcgaacaaacctctctcgaggcccccagtttccgtccctacacggagct 1003
 Db 374 TCTTCCAGACCAACCTCTCTCGAGGCCCTTAACTTTCCTCCCTACACAGGAGCT 433
 Qy 1004 cctccccaaggtatagtcgacgcgttatcgtgctatagcatatgtctcccaaaaa 1063
 Db 434 CTTCCCAAGGATGATGACGACCGTTATCTGCTTATGACATTAATGCTCCCAAAAA 493
 Qy 1064 aaactctgtcctgcatacctcgtgtatcaaatgacattttaaccaatccaatcaaaa 1123
 Db 494 AAACCTCTTGGCTGCACTGCTGTATACATGATGACATTTTAAACCAATCCATCTAATA 553
 Qy 1124 tctgcacgaatccacccctgtgtcccggaatcgtgttggttctcttcttctacacgcag 1183
 Db 554 TGTGCCAGAAATCCACCTGTGGCCGGAATCGTTTGGTCTCTTCTACTCCATGCGAC 613
 Qy 1184 a 1184
 Db 614 A 614

RESULT 5
 BE387718
 LOCUS BE387718 638 bp mRNA EST 21-JUL-2000
 DEFINITION 601267377F1 NIH_MGC_44 Homo sapiens cDNA clone IMAGE:3621618 5',
 mRNA sequence.
 BE387718
 ACCESSION BE387718 GI:9333083
 VERSION BE387718
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 REFERENCE 1 (bases 1 to 638)
 NIH-MGC <http://www.ncbi.nlm.nih.gov/MGC/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert_Strausberg@nih.gov
 Tissue Procurement: ATCC
 cDNA Library Preparation: Ling Hong/Rubin Laboratory
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: L1CM295 row: j column: 19
 High quality sequence stop: 634.
 Location/Qualifiers
 1. 638
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:3621618"
 /clone_lib="NIH_MGC_44"
 /tissue_type="endometrium, adenocarcinoma cell line"
 /lab_host="PH10B (phage-resistant)"
 /note="Organ: uterus; Vector: pOTB7; Site_1: XhoI; Site_2:
 EcoRI; cDNA made by oligo-dT priming. Directionally
 cloned into EcoRI/XhoI sites using the following 5'
 adaptor: GGCACGAG(G). Library constructed by Ling Hong
 in the laboratory of Gerald M. Rubin (University of
 California, Berkeley) using ZAP-cDNA synthesis kit
 (Stratagene) and Superscript II RT (Life Technologies)."
 BASE COUNT 160 a 159 c 141 g 178 t
 ORIGIN

Oy	859	ggtgcctatcggagacgagcgactccgaggtccaagtccgaacgagcaggaccatcaat	918
Dd	751	gggtgctaattctggagcca--ggatccgagttccaggttcagacagggaag---ccatcaaat	805
Oy	919	gccaccgagaagatbtctctcccatcttccgaaccaacaacctctcgagggcccta	977
Dd	806	ggaccggaaagatgtttcttcatacttccgaaccaaacctctcgagggcccta	864
RESULT	7		
BE869679		BE869679	EST
LOCUS		601445783.p1.NH_MGC_65 Homo sapiens CDNA clone IMAGE:3849620 5'	27-SEP-2000
DEFINITION		mRNA sequence.	
ACCESSION		BE869679	
VERSION		BE869679.1	
KEYWORDS		GT:10318364	
SOURCE		human.	
ORGANISM		Homo sapiens	
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
AUTHORS		Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	
TITLE		1 (bases 1 to 699)	
JOURNAL		NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/ National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999)	
COMMENT		Contact: Robert Strausberg, Ph.D. Tel: (301) 496-1550 Email: Robert_Strausberg@nih.gov Tissue Procurement: ATCC CDNA Library Preparation: Life Technologies, Inc. CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: LILCM554 row: f column: 21 High quality sequence stop: 690.	
FEATURES		location/Qualifiers	
source		1..699	
		/organism="Homo sapiens"	
		/db_xref="taxon:9606"	
		/clone="IMAGE:3849620"	
		/clone_lib="NIH MGC 65"	
		/tissue_type="adenocarcinoma"	
		/lab_host="DH10B (phage-resistant)"	
		/note="Organ: colon; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.8 kb. Library constructed by Life Technologies."	
BASE COUNT		218 a 117 c 198 g 166 t	
ORIGIN			
Query Match		20.6%; Score 588.2; DB 136; Length 699;	
Best Local Similarity		95.4%; Pred. No. 5.5e-117;	
Matches 669; Conservative 1; Mismatches 24; Indels 7; Gaps 6			

OY	248	aaaagagttgttgcttcgagcagaactatgaatacgccgcaaaaggacaagaagt	307
Db	241	AAAAAGGTTGGTCTGTGAGCGCAACTGTGAATGAAAGGGCCAAAAGAGCGCAAGAAGAT	300
OY	308	gggagagaatttcgaaaaaccttgaagatccagaagaatgcgtccagaagagtttatgaccctc	367
Db	301	GGGAGAAAGTTTCGAAAACCTTGAAAGATCCAGAGAAATGTCCAGAGAGGTTTTATGACCCTC	360
OY	368	gatctctatatgaaggtctacagagacagaagaagacagacagcagagtagtacaggaac	427
Db	361	GATCTCTATATGAAAGGCTACAGGAACAAGAGACAGAGAGCAGCAGAGAGTAGAGAGAAC	420
OY	428	agttcaaatccaatacatgtaagaagcgttgaatyagaatgagaccaacttcctgtagt	487
Db	421	AGTTCAAAATTCAAAAACAATGATAGAGCGCTTGATGATGAATGATGAGACCAACTTCCTTGATG	480
OY	488	agttcttcgacagaagaagacactaatagaagaacaagaagaagaagaagacttgaagaac	547
Db	481	AGCTTTCTGCACAGCAGGAACTATTBGAAGAACGA-GAAGAGAGAGAGAA-TGAAGAGAC	538
OY	548	tgaaggaatcacagaataaacctccaagaaggttggaaatttcctaagaagaacaagaagaag	607
Db	539	TGAAGGATATACAGAAATACTCAAGAAAGGTGGGAATTTCTCAGAGAGAAACAGAGGGAAG	598
OY	608	tggaaaagaactgtgactgtgaagcctatagaaccaagaacaagtctcccagcgaagc	667
Db	599	GGGAAAGAGAA-TGACTGTGAAGGCTATTGGAAC--AGCAAAAAGTTCTCCAGCGCACAGC	655
OY	668	tgt-tggcagagagctgtgaaag-cataaagagccagaagaatg	706
Db	656	TGTGGGCGCAGAGCTGTGAAGCCATPAGAGCTCAGAGAGTG	696

RESULT	8		
LOCUS	BE392702	664 bp	mRNA EST 21-JUL-2000
DEFINITION	601307895F1 NIH_MGC_44 Homo sapiens CDNA clone IMAGE:3626302 5'		
ACCESSION	BE392702		mRNA sequence.
VERSION	BE392702.1	GI:9338166	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
TITLE	1 (bases 1 to 664)		
JOURNAL	NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/		
COMMENT	National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999) Contact: Robert Strausberg, Ph.D. Tel.: (301) 496-1550 Email: Robert.Strausberg@nih.gov Tissue Procurement: ATCC CDNA Library Preparation: Ling Hong/Rubin Laboratory CDNA Library Arrayed by: The I.M.A.G.E. Consortium (NLN) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/NLNL at: http://image.llnl.gov Plate: LNCM307 row: m column: 23 High quality sequence stop: 631. Location/Qualifiers 1..664 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:3626302" /clone_lib="NIH_MGC_44" /tissue_type="endometrium, adenocarcinoma cell line" /lab_host="DH10B (phage-resistant)" /note="Organ: uterus; Vector: pORF7; Site:1; XhoI; Site_2 EcoRI; cdna made by oligo-dt priming. Directionally cloned into EcoRI/XhoI sites using the following adaptor: GGCACGAG(G). Library constructed by Lin Hong"		

in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-CDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."

BASE COUNT 158 a 176 c 148 g 182 t

Query Match 20.5%; Score 583.4; DB 107; Length 664;
Best Local Similarity 98.4%; Pred. No. 5.9e-116;
Matches 631; Conservative 0; Mismatches 6; Indels 4; Gaps 4;

914 tcaatgccacgggaagattgtctctccatcttcggaacacacctctctcgagccc 973
1 tcaatgccacgggaagattgtctctccatcttcggaacacacctctctcgagccc 60
974 cctagttctccgtccctcagacaggagctctccccaagggtatagcagcgtcatg 1033
61 cctagttctccgtccctcagacaggagctctccccaagggtatagcagcgtcatg 120
1034 ctgcctataggacatgatgctcccaaaaaaacctcttgcctcgtcatcctgtgacac 1093
121 ctgcctataggacatgatgctcccaaaaaaacctcttgcctcgtcatcctgtgacac 180
1094 atgacattttaaacaatccaatcaataaagtgcagaaatccacgtgtgacagatcg 1153
181 atgacattttaaacaatccaatcaataaagtgcagaaatccacgtgtgacagatcg 240
1154 tcttggctctctcttctacacacacacacacacacacacacacacacacacacac 1213
241 tcttggctctctcttctacacacacacacacacacacacacacacacacacacac 300
1214 tcaactgatatgag 1273
301 tcaactgatatgag 360
1274 aatagacacggcgctgcttgcacaaagggttgggtttatgtcttcttcttctt 1333
361 aatagacacggcgctgcttgcacaaagggttgggtttatgtcttcttcttctt 420
1334 cccgac 1393
421 cccgac 480
1394 taa-atgaatgt 1452
481 taaatgt 540
1453 cccggtatgataatgagagagagagagagagagagagagagagagagagagag 1511
541 cccggtatgataatgagagagagagagagagagagagagagagagagagagag 600
1512 aagtaaatgataaac 1552
601 tagt-aagatgataaac 639

RESULT 9
BE180639/c 631 bp mRNA EST 22-JUN-2000
LOCUS BE180639
DEFINITION RC3-HR0625-040500-022-d11 HR0625 Homo sapiens cDNA, mRNA sequence.
ACCESSION BE180639
VERSION BE180639.1 GI:8659815
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 631)
Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R., Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordino, S., Costa, F.F., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H., Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and

TITLE Simpson, A.J.
JOURN. Shotgun sequencing of the human transcriptome with ORF expressed
MEDLINE sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?lt=612-RC3-HR0625-040
500-022-d11&f3=2000-05-04&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 5
High quality sequence stop: 630.
Location/Qualifiers
1. 631

FEATURES
source
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_id="HR0625"
/dev_stage="Adult"
/note="Organ: head,neck; Vector: puc18; Site:1: Smal;
Site:2: Smal; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."

BASE COUNT 127 a 145 c 182 g 177 t

Query Match 20.4%; Score 581.8; DB 105; Length 631;
Best Local Similarity 98.7%; Pred. No. 1.3e-115;
Matches 597; Conservative 0; Mismatches 7; Indels 1; Gaps 1;

542 aagaactgaaggaatcacagaataaacctcaagaaggttgaattctcagaagaaca 601
631 aagaactgaaggaatcacagaataaacctcaagaaggttgaattctcagaagaaca 572
602 aggaagtgaaggaatcacagaatgagagcctatagaacaagaagaagttccacag 661
571 aggaagtgaaggaatcacagaatgagagcctatagaacaagaagaagttccacag 512
662 cgaagctgttgcagagagctgtgaagcatalaagagctcagagagtgcaacagtg 721
511 cgaagctgttgcagagagctgtgaagcatalaagagctcagagagtgcaacagtg 452
722 gactgaacacggagacccctgagcagatgacagaatcaagaagcctcctcgaagtc 781
451 gactgaacacggagacccctgagcagatgacagaatcaagaagcctcctcgaagtc 392
782 tcggaacacccctcctgagtgagccctcctcctcctcctcctcctcctcctcctc 841
391 tcggaacacccctcctgagtgagccctcctcctcctcctcctcctcctcctcctc 332
842 gcatcctccagagcctgagtgagcctcctcctcctcctcctcctcctcctcctcctc 901
331 gcatcctccagagcctgagtgagcctcctcctcctcctcctcctcctcctcctcctc 272
902 gcgaagcagacatcaatgacacacgggaagattgtctcctcctcctcctcctcctc 961
271 gcgaagcagacatcaatgacacacgggaagattgtctcctcctcctcctcctcctc 212
962 tctctgagagccctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 1021
211 tctctgagagccctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 152

QY	1022	gagccctcatcgtgcctcgtatgacatattgctccctcaaaaaaactccttgcgcgat	1081
Db	151	GGACCGTTATGCTGCTGCTTAGGCAATTATGCTCCCTTAAAAAAAACCTCTTTCCTGTCAT	92
QY	1082	ccctgtgtacaacatgacatltttaaccaatccaatcctaaatgctgcagaatccactg	1141
Db	91	CGTGTCACACATGACATTTTAAACCATGCCAATGTAAATAATGTGCCAG-ATCCACCTG	33
QY	1142	tgagcc	1146
Db	32	TCGCC	28
RESULT	10		
LOCUS	BE736929		
DEFINITION	BE736929	895 bp	mRNA
ACCESSION	601306864	NIH_MGC_39	Homo sapiens cDNA clone IMAGE:3641343 5', mRNA sequence.
VERSION	BE736929		
KEYWORDS	BE736929.1	GI:10150832	
SOURCE	EST.		
ORGANISM	human.		
REFERENCE	Homo sapiens		
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.		
TITLE	1 (bases 1 to 895)		
JOURNAL	NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/ .		
COMMENT	National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999)		
	Contact: Robert Strausberg, Ph.D.		
	Tel: (301) 496-1550		
	Email: Robert.Strausberg@nih.gov		
	Tissue Procurement: ATCC		
	cDNA Library Preparation: Ling Hong/Rubin Laboratory		
	cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)		
	DNA Sequencing by: Incyte Genomics, Inc.		
	Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:		
	http://image.llnl.gov		
	Plate: LUCM346	row: P	column: 16
	High quality sequence stop: 726.		
FEATURES			
source	Location/Qualifiers		
	1..895		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone="IMAGE:3641343"		
	/clone_1lb="NIH_MGC_39"		
	/tissue_type="adenocarcinoma"		
	/lab_host="DH10B (phage-resistant)"		
	/note="Organ: pancreas; Vector: pORF7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)"		
BASE COUNT	232 a	183 c	218 g 262 t
ORIGIN			
Query Match	20.2%	Score 575.2;	DB 135; Length 895;
Best Local Similarity	91.9%;	Pred. NO. 3.7e-114;	
Matches 741; Conservative	0;	Mismatches 46;	Indels 17; Gaps 12
QY	1170	ctactccactgcagatgacaaacctgtccgctgcacatttccctcactgcgatattggag	1229
Db	1	CTACTCCACTGCAGATGACAAACCTGTCGCCCTG-CACCTTCCTCAGATGATA-TGGGAG	58
QY	1230	gagggcgaagcccgacgcgaagtccactcaaaaatgcccgagagatagagccgagctg	1289
Db	59	GAGGGCAAGGCCACACCAAGTTCCATTAATAAATCCCCGAGAGATAGCACCGGCTGG	118

QY	1290	cttgcgaagaagggttggtggtttatgtcttctggtttttctttcttccgcagacgcaaga	1349
Db	119	ctgcgaagaagg--ttgggggtttatgcttgcgtttttttttttttccgcacacacgaaga	175
QY	1350	agtaaggagagatattatgagcagggtgtatttaaacatctcatctgttaaatgaatgtgtg	1409
Db	176	acttaaggagagttta-tggacagatgtgtattttaaacattcta-tgtnaatgaattgctt--g	231
QY	1410	ttgtgtctactgtcatctgtggaagcatgcyggggaaagaaacatgaccaglaatgaatgg	1469
Db	232	ttgggtttctactgcattgtgtgagacatgccccggggaagaaagacatgaccagtaatgaatgg	291
QY	1470	agcccttcctcctggaactaaaccgctctgtatgttgytgcactaaagtaagaatgaacc	1529
Db	292	acccttctcccttggaactaaaccgctctgtatgttgytgcactaaagtaagaatgaatgaacc	351
QY	1530	ccatctgtcgggggtgtgtaacttcgaactcggatgcattgtgaaagcttccatccctt	1589
Db	352	ccatctgtcgggggggtgtgtaacttcgaactcggatgcattgtgaaagcttccatccctt	411
QY	1590	ggccatctccctctctcctctcctcctccaaaccattatgcaagaaagagctgctaacaag	1649
Db	412	ggccatctccctctcctcctcctcctccaaaccattatgcaagaaagagagctgctaacaag	470
QY	1650	aagcgttcacatcacaacctttctctgcctcctggaataattatgttgttttga	1709
Db	471	aagcgttccatctcaaacctttttctcctcctcctggaataattatgttgttttga	528
QY	1710	taaagatttagtttaagatcttaattttaagaaacaaagctgagcttctttactaa	1769
Db	529	taaagatttagttttaagatcttaattttaagaaacaaagctgagcttctttactaa	588
QY	1770	tagcgaagacatcagaacatgcaagtagtatgttaatgagatgtaactattcttcgagcct	1829
Db	589	tagcgaagacatcagaacatg--agtgtagtatgttaatgagatgtaactattcttcgagcct	647
QY	1830	ctggaatcctcaatattgttaaatga--gtggagacacactgtcatatgtgacatctcat	1887
Db	648	ctggaatcctcaatattgttaaatgaagacgtggagacacacactgtcatatgtgacatctcat	707
QY	1888	gagggcctctcctgtttaaagcatatatactgtgctttaaacttgatgacatattcta	1947
Db	708	gagggcctctcctgtttaaagcatatatactgtgctttaaacttgatgacatattcta	765
QY	1948	acctaaagtgctgcctcctgacttctt	1973
Db	766	cctaagctgctgcctcctgacttcttctt	791
RESULT	11		
	AM327476		
LOCUS	568 bp	mrna	EST
DEFINITION	gdc3g909.x1 NIH_MGC_2 Homo sapiens cDNA clone IMAGE:2846800 5', mRNA		
ACCESSION	AM327476		
VERSION	AM327476.1		
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
TITLE	1 (bases 1 to 568)		
JOURNAL	NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.		
COMMENT	National Institutes of Health, Mammalian Gene Collection (MGC)		
	Unpublished (1999)		
	Contact: Robert Strausberg, Ph.D.		
	Tel: (301) 496-1550		
	Email: Robert_Strausberg@nih.gov		
	Tissue Procurement: ATCC		
	cDNA Library Preparation: Edge Biosystems		
	cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMD)		
	DNA Sequencing by: NIH Intramural Sequencing Center (NISC)		
	Clone distribution: MGC clone distribution information can be		

found through the I.M.A.G.E. Consortium/LNL at:
www.bio.lnl.gov/bbrp/image/image.html

Plate: LINC0028 row: N column: 17
Seq primer: -21M13 forward primer (ABI).

FEATURES

Source

Location/Qualifiers
1..568

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2846800"
/clone_lib="NTH_MGC_2"
/tissue_type="T cell leukemia"
/cell_line="MGC2"
/note="Organ: Blood; Vector: pOTB7a; Library prepared by
Edge Biosystems."

BASE COUNT 164 a 165 c 123 g 116 t
ORIGIN

Query Match 19.8%; Score 564.8; DB 89; Length 568;

Best Local Similarity 99.6%; Pred. No. 6.1e-112;

Matches 566; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
OY 558 cagaataacctcaagaaggttggaatttcagaagaagaagaagtgaaagaa 617
Db 1 CAGAAATACCTCAAGAAAGTTGGAATTTCTCAAGAAACAAGAAAGTGAAGAA 60
OY 618 actgactgtgaagctatagaaccaagaagaagctcccgagcgaagctgtgcaag 677
Db 61 ACTGACTGTGATCTCTATAGAAACCAAGACAGAGTTCTCCGAGGAGACTTTGGCAGG 120
OY 678 agctgtgaagcaataagactcagaagatgtgcaacagtttgaagaactgaacccgagacc 737
Db 121 AGCTGTGAAGCTAAGACTCAGAGAGTGGCAAGTGTGAAGAACTGAAGACCGAGACC 180
OY 738 tgaagccagatgaacaagaatcaagaagctcactcgaagctcgcgaagaacactccct 797
Db 181 TGAGCCAGATGACAAAGAAATCAAGAGCCCTCATCTGCAAGTCTCTGGAAACACTCTCCT 240
OY 798 gagtggccctccatcacactgcccctctgtcgaatgtatcgcgacatccctccagacct 857
Db 241 GAGTGGCCCTCCATCCACTGCGCTCTGCTGCAATGATGATGCGGCAATCTCCAGGCT 300
OY 858 ggggtgctactcttggaagcagcagactccgagtcagctcagaacagcgaagcacaatcaa 917
Db 301 GGGTCCCTACTCTGGAGCAGAGCACTCCGAGTCCAGCTCAGACAGCAAGCAGCATCAA 360
OY 918 tggccacggaaagatgtgtctctcatcttcgaacaacacctccctccgagccccccta 977
Db 361 TGCCACCGGAAAGATTGTCTCTCATCTTCGGAACCAACACCTTCTCGAGGCCCTCA 420
OY 978 gtcttcgcgtccctacacaggaagctcctcccaaggtagatcgagccgttcacgtctgc 1037
Db 421 GTTCTCGCTCCCTACACAGGAGCTCTCCCAAGGGTAAATCGGACCTTCATGCTGC 480
OY 1038 ctatagcattatgtccctcaaaaaaaactccttgcctgcataccctgtgtacaacatga 1097
Db 481 CTATGAGCATTAATGTCTCCCAAAAAAACTCTTGGCTGCATCTGTGTACACATGA 540
OY 1098 cattttaaccaatcattcaaaaatg 1125
Db 541 CATTTTAAACCAATCAATCTAAATATG 568
```

RESULT 12
AM592821/c 554 bp mRNA EST 22-MAR-2000

LOCUS AM592821 554 bp mRNA EST 22-MAR-2000

DEFINITION hg6601.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone

IMAGE:2944800 3', mRNA sequence.

ACCESSION AM592821

VERSION AM592821.1 GI:7280012

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 554)

NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index

Unpublished (1997)

Contact: Robert Strausberg, Ph.D.

Tel: (301) 496-1550

Email: Robert.Strausberg@nih.gov
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (infoimage.lnl.gov) for further information.

Seq primer: -400P from Glibco
High quality sequence stop: 438.

FEATURES

Source

Location/Qualifiers
1..554

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2944800"
/clone_lib="Soares_NFL_T_GBC_S1"
/lab_host="DH10B"

/note="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with
a modified polylinker; Site: 1: Not I; Site: 2: Eco RI;
Equal amounts of plasmid DNA from three normalized
libraries (fetal lung NDHL19W, testis NHT, and B-cell
NCI-GAP-GCB1) were mixed, and ss circles were made in
vitro. Following HAP purification, this DNA was used as
tracer in a subtractive hybridization reaction. The driver
was PCR-amplified cDNAs from pools of 5,000 clones made
from the same 3 libraries. The pools consisted of
I.M.A.G.E. clones 297480-302087, 682632-687239,
726408-728711, and 729096-731399. Subtraction by Bento
Soares and M. Fatima Bonalao.

BASE COUNT 151 a 122 c 135 g 146 t
ORIGIN

Query Match 19.4%; Score 552.4; DB 92; Length 554;

Best Local Similarity 99.8%; Pred. No. 2.9e-109;

Matches 553; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
OY 2249 tgggtgtgagactgagttagacacacacaaatgtgttgaacactgtgatccgcgaagca 2308
Db 554 TGGGTGTGAGACTGAGTGAACACACACAAATGTTGACACTGTGATGCCGCGAGCA 495
OY 2309 gagcagctactgactttaacatggcagagagcccttgatctcacagccacccc 2368
Db 494 GAGCAGCTACTGACTTTGAACATGGGCAAGAGGCCCTGGATCTCATCAAGCCACTGC 435
OY 2369 ttltccctccagtagacagctcgtgtgcccattgagcagatgagcactccctga 2428
Db 434 TTTTCCCTTCCAGTAGACGTGACACTGTGTCGCCATTGGCAGATGGGACTTCCTCA 375
OY 2429 ccataactgatgcttgtgaattctccctcccttccagaactactcgtgtcaatgtt 2488
Db 374 CCCATTAAGTGAATGCTTGTGAATTTCTCTCTTTCGAACTACTCTGTCTAATTTGT 315
OY 2489 ctgcaagatgagggcagcatcacaactcactgaacaagaacattagtaaaacttgc 2548
Db 314 CTGCCAGTAGGCGCCATCAGCTCCATCTCTGACAAACAGACATTTAGGTAAACTTTGT 255
OY 2549 aggcacctctgctctctgcttcatgttctcgtgtatagctcgtgttlaaacagat 2608
Db 254 AGGCACCTTCTGCTCTCTGCTTCAATGCTTCCTGATGATGCTGTGTTATTACAGCAT 195
OY 2609 gtacccaaagagctcacatgtttaacagagggcagccagagacataaagtcatcact 2668
Db 194 GTACCCAAAGAGCTCACATTTGTAAGAGGCGAGGCGAGGACATCAAAAGTCAACATCT 135
OY 2669 ttatgtgcatgactctaaagagcattactgatactcagagcctctgtgtgaaag 2728
Db 134 TTATGTGCATGACTTTAAGAGGCCATTACTGTAATCTCATGAGCCCTTGTGATGGAAG 75
```

OY	2729	aagtgcagcagggtgctcgggtttaaacacccatcaataaccatgaagaattaaacctt	2788
DB	74	AAGTTGACAGAGGGCTTTCAGCGGTAAAAACAATCCATTACTGAAGCTAATAAACTTG	15
OY	2789	tcaagacaacaaaaa	2802
DB	14	TCAGAGAACAACGAA	1
RESULT	13		
LOCUS	A1205174/c		
DEFINITION	ap13a11.x1 Schiller oligodendrogloma Homo sapiens cDNA clone IMAGE:1955228 3' similar to W:C25A1.1 CE08366 ;, mRNA sequence.		
ACCESSION	A1205174		
VERSION	A1205174.1		
KEYWORDS	EST.		
SOURCE	Homo sapiens		
ORGANISM	Homo sapiens		
REFERENCE	Mammalia; Eutheria; Primates; Catarrhini; Homiidea; Homo.		
AUTHORS	I (bases 1 to 568)		
TITLE	Kilzman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Maira,M., Martin		
JOURNAL	,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B.,		
COMMENT	White,Y., Wyllie,T., Waterston,R. and Wilson,R. Washu-NCI human EST project Unpublished (1997) Contact: Wilson RK Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: est@wustl.edu This clone is available royalty-free through LINL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information. Possible reversed clone: similarity on wrong strand Seq primer: -40UP from Gibco High quality sequence stop: 449. Location/Qualifiers 1..568 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:1955228" /clone_lib="Schiller oligodendrogloma" /sex="male" /tissue_type="Oligodendrogloma" /dev_stage="44 years" /lab_host="SOLR" /note="Organ: brain; Vector: pBluescript SK- (Stratagene); Site_1: EcoRI, Site_2: XhoI; Double-stranded cDNA was prepared from human oligodendrogloma using primer 5'-GAGGAGAGAGAGAGAGAGAACTGCTGACT(18)-3'. An EcoRI adaptor was used on the 5' end of the cDNA as follows: 5'-AATTCGCCAGCAG-3'. The library was size-selected and size is 1.7 kb, with a range from 0.4-12 kb. Tumor identification by consensus pathology; contains chromosome 1p and 19q deletion as determined by CGH. This library was constructed by Dr. Martin Schiller (Johns Hopkins University)."		
BASE COUNT	96 a	147 c	135 g
ORIGIN	147 c	135 g	190 t
Query Match	18.8%; Score 534.4;	DB 17;	Length 568;
Best Local Similarity	96.3%; Pred. No. 2.3e-105;		
Matches 547; Conservative	0; Mismatches 21;	Indels	0; Gaps 0;
OY	376	tatgaagctcacagcagaacagcagcagcagcagcagtgatcagcaggaacagttcaaa	435
DB	568	TATGGAAGGCTACAGAACAGAGACAGCTCACGACGAGACTACGAGAACCATTAACAAA	509

Oy	436	tttaaaaaatcgtgtgaaggagccttagaataagaatgtgagccaacttctcttgatbaagttct	495
Oy	436	tttaaaaaatcgtgtgaaggagccttagaataagaatgtgagccaacttctcttgatbaagttct	495
Db	508	TCACATCATGTGTGTAAGAGGCTTACATGTAAGTGAAGACGAACCTTCTTGATGAGGTTTCT	449
Oy	436	cgacgcaggaactaataatgaanaagcaacgaagaagaagaactctgaagaacttgaaggaa	555
Db	448	CGACAGCAGGAACCTATATGAAAGCAGACGACGAAGAAGTAAGAAGTGAAGCA	389
Oy	556	tacagaataaacttcaagaaggcttgaaattcttcaagaagaacagaagaagtgtgaaag	615
Db	388	TACACCATTAACCTTCAGAAAGGTTGGAATTTCTCTCAAGAGAACACGAAGAGTGGAAAG	329
Oy	616	aaacgactgtggaagccctatagaaacccaagaacaaagttctcccaagcgaagctgttgca	675
Db	328	AAGCTGACTGTGTAAACCCCTATAGAACCCAGAAACAAGTTCTCTCAGCCGAAAGCTGTGGCA	269
Oy	676	ggaagctgtgaaacaaagaagctcagaagatgtgcaacagctgtgaagaagactgaaccggac	735
Db	268	GGAGCTGTGTAAACCTATATGAAAGCCTCAGAGAGTGGCAACAGTGTAAAAGACTGAAGCCGAC	209
Oy	736	cctgagccagatgacgaagaatcaagaagccctatcctgcaagttcttcggaacaacctcc	795
Db	208	CCTGAGCCAGATGACAAAGATCAAGAGCCCTCATCTCTGCAAGTCTCTCGAAGCAACCTCC	149
Oy	796	ctgagtgagccctccatccactgagccctctgtgtgagatgatatgtgcattctctccagc	855
Db	148	CTGAATGTGCCCCCTCATCCACCTGCCCCCTCTGTCGAGATGTATGTGGCATCTCCAGGC	89
Oy	856	ctgagtgagctactctcggagacagagaccccgagtcagctcaagacagcgaagacacatc	915
Db	88	CTGGGTGCTTACTCTCTGGAGCAGCCACCTCCGAGTCCAGCTCAGACAGCGAAGGCACATC	29
Oy	916	aatgccaccggaagaatgtgtctctcca	943
Db	28	AATACCACCCCAAGATGTGTCTCTCA	1
RESULT	14		
LOCUS	AI031969/c		
DEFINITION	ow55d01.x1 Soares parathyroid tumor_NbHRA Homo sapiens cDNA clone		
ACCESSION	AI031969		
VERSION	AI031969.1	GI:3250181	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS	1 (bases 1 to 577)		
TITLE	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap. National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index		
JOURNAL	Unpublished (1997)		
COMMENT	Contact: Robert Strausberg, Ph.D. Tel: (301) 496-1550 Email: Robert.Strausberg@nih.gov cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo Ph.D.		
FEATURES	cDNA Library Arrayed by: Greg Lennon, Ph.D. DNA Sequencing by: Washington University Genome Sequencing Center Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html Insert length: 1621 Std Error: 0.00 Seq primer: -40m13 fwd: ET from Amersham High quality sequence stop: 444. Location/Qualifiers		
source	1..577		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone="IMAGE:1650721"		
	/clone_lib="Soares_parathyroid_tumor_NbHRA"		

Thu Mar 1 08:56:57 2001

us-09-602-597-1.rst

Page 13

OY	1627	tgcaggaaagcgtctctaacaagaacgcttc	1657
Db	540	TGCACGGAAGACTGCTTAACAAGAAGCGTTC	570

```
Search completed: February 28, 2001, 23:43:20
Job time: 19386 sec
```

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 28, 2001, 22:33:49 ; Search time 115.04 Seconds
(without alignments)
3992.584 Million cell updates/sec

Title: US-09-602-597-1

Perfect score: 2850
Sequence: 1 aagccctgaaggggtcctaaag.....aaaaaaaaaaaaaaaaaaaaa 2850

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 280836 seqs, 80580151 residues

Total number of hits satisfying chosen parameters: 561672

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_NM:*
1: /cgn2_6/ptodata/1/ina/5A.COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B.COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6.COMB.seq:*
4: /cgn2_6/ptodata/1/ina/PCtUS.COMB.seq:*
5: /cgn2_6/ptodata/1/ina/backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	92	3.2	7218	1	US-08-232-463-14 Sequence 14, Appl
2	67.6	2.4	1307	2	US-08-960-022-17 Sequence 17, Appl
3	67.2	2.4	1641	1	US-08-300-903A-8 Sequence 8, Appl
4	67	2.4	2082	2	US-08-785-310A-2 Sequence 2, Appl
5	65.4	2.3	6671	1	US-08-280-443-1 Sequence 1, Appl
6	65.4	2.3	6671	1	US-08-457-459-1 Sequence 1, Appl
7	65.4	2.3	6671	1	US-08-535-678-1 Sequence 1, Appl
8	65.4	2.3	6671	4	PCT-US95-02275-1 Sequence 1, Appl
9	64.8	2.3	144	1	US-08-702-344-26 Sequence 26, Appl
10	64.8	2.3	1147	1	US-08-665-716-1 Sequence 1, Appl
11	64.8	2.3	1172	1	US-07-945-288-9 Sequence 9, Appl
12	64.8	2.3	1172	1	US-08-462-831-9 Sequence 9, Appl
13	64.8	2.3	1172	1	US-08-461-809-9 Sequence 9, Appl
14	64.8	2.3	1172	1	US-08-461-441-9 Sequence 9, Appl
15	64.8	2.3	1172	4	PCT-US93-08518-9 Sequence 9, Appl
16	64	2.2	1817	1	US-08-473-981A-5 Sequence 5, Appl
17	64	2.2	1817	2	US-08-474-087-5 Sequence 5, Appl
18	64	2.2	2381	1	US-08-021-608D-9 Sequence 9, Appl
19	64	2.2	2381	1	US-08-726-160-9 Sequence 9, Appl
20	64	2.2	2381	4	PCT-US94-01782-9 Sequence 9, Appl
21	64	2.2	2384	1	US-08-021-608D-1 Sequence 1, Appl
22	64	2.2	2384	1	US-08-726-160-1 Sequence 1, Appl
23	64	2.2	2384	4	PCT-US94-01782-1 Sequence 1, Appl
24	63.8	2.2	240	1	US-08-628-417-6 Sequence 6, Appl
25	63.8	2.2	5173	1	US-08-242-677-1 Sequence 1, Appl
26	63.6	2.2	140	1	US-08-628-417-5 Sequence 5, Appl
27	63.6	2.2	630	1	US-08-185-414E-1 Sequence 1, Appl
28	63.6	2.2	2233	1	US-08-496-631-1 Sequence 1, Appl

29	63.4	2.2	2447	2	US-09-014-969-14 Sequence 14, Appl
30	63.4	2.2	2744	3	US-09-071-101-1 Sequence 1, Appl
31	63.4	2.2	2744	3	US-09-369-618-1 Sequence 1, Appl
32	63.4	2.2	2744	3	US-09-369-617-1 Sequence 1, Appl
33	63.2	2.2	1882	2	US-09-370-253-1 Sequence 1, Appl
34	63.2	2.2	2550	5	5258287-23 Patent No. 5258287
35	63	2.2	991	3	US-08-520-678A-25 Sequence 25, Appl
36	63	2.2	991	3	US-08-924-747-25 Sequence 25, Appl
37	63	2.2	3527	2	US-08-909-965C-7 Sequence 7, Appl
38	62.8	2.2	1066	1	US-08-157-101A-4 Sequence 4, Appl
39	62.8	2.2	1098	3	US-09-248-335-35 Sequence 35, Appl
40	62.8	2.2	1134	3	US-09-248-335-29 Sequence 29, Appl
41	62.6	2.2	2852	3	US-09-027-137-2 Sequence 2, Appl
42	62.4	2.2	1461	3	US-08-722-126A-4 Sequence 4, Appl
43	62.4	2.2	1461	4	PCT-US95-04258-4 Sequence 4, Appl
44	62.4	2.2	1738	2	US-08-379-482A-2 Sequence 2, Appl
45	62.2	2.2	688	5	5498694-3 Patent No. 5498694

ALIGNMENTS

RESULT 1
US-08-232-463-14/C
Sequence 14, Application US/08232463
Patent No. 5670367
GENERAL INFORMATION:
APPLICANT: DORNER, F.
APPLICANT: SCHEIFLINGER, F.
APPLICANT: FALKNER, F. G.
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 1800 Diagonal Road, Suite 500
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/232,463
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/935,313
FILING DATE:
APPLICATION NUMBER: EP 91 114 300.6
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30472/114 IMMU
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)836-9300
TELEFAX: (703)683-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: pTZ9pC-Fls
US-08-232-463-14
Query Match 3.2%; Score 92; DB 1; Length 7218;

Query Match	2.3%	Score 64.8	DB 1	Length 1147
Best Local Similarity	68.2%	Pred. No. 5.8e-06		
Matches 90; Conservative	0	Mismatches 42	Indels 0	Gaps 0

[illegible][illegible]

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2001, 17:05:06 ; Search time 16.5 seconds

(without alignments)
526.377 Million cell updates/sec

Title: US-09-602-597-2

Perfect score: 1306

Sequence: 1 MDGGDDCNLIKRRFVSEAE.....NATGKIYSSIFRTFTLEAP 254

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 268485 seqs, 34193795 residues

Total number of hits satisfying chosen parameters: 268485

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
A.Geneseq.36.*
1: /SIDSL/gcgdata/geneseq/AA1980.DAT.*
2: /SIDSL/gcgdata/geneseq/AA1981.DAT.*
3: /SIDSL/gcgdata/geneseq/AA1982.DAT.*
4: /SIDSL/gcgdata/geneseq/AA1983.DAT.*
5: /SIDSL/gcgdata/geneseq/AA1984.DAT.*
6: /SIDSL/gcgdata/geneseq/AA1985.DAT.*
7: /SIDSL/gcgdata/geneseq/AA1986.DAT.*
8: /SIDSL/gcgdata/geneseq/AA1987.DAT.*
9: /SIDSL/gcgdata/geneseq/AA1988.DAT.*
10: /SIDSL/gcgdata/geneseq/AA1989.DAT.*
11: /SIDSL/gcgdata/geneseq/AA1990.DAT.*
12: /SIDSL/gcgdata/geneseq/AA1991.DAT.*
13: /SIDSL/gcgdata/geneseq/AA1992.DAT.*
14: /SIDSL/gcgdata/geneseq/AA1993.DAT.*
15: /SIDSL/gcgdata/geneseq/AA1994.DAT.*
16: /SIDSL/gcgdata/geneseq/AA1995.DAT.*
17: /SIDSL/gcgdata/geneseq/AA1996.DAT.*
18: /SIDSL/gcgdata/geneseq/AA1997.DAT.*
19: /SIDSL/gcgdata/geneseq/AA1998.DAT.*
20: /SIDSL/gcgdata/geneseq/AA1999.DAT.*
21: /SIDSL/gcgdata/geneseq/AA2000.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	779	59.6	168	20 Y48610	Human breast tumor
2	143	10.9	1135	21 Y68784	Human acid sequenc
3	142.5	10.9	1178	18 W50763	Mannose-1-phosphat
4	141.5	10.8	562	16 R70491	Leucocytotoxin prot
5	141	10.8	1239	20 Y55931	Human ZC1 protein.
6	139	10.6	1233	20 Y55954	Mouse STRE20-relate
7	133	10.2	1162	21 Y58500	HHV8 ORF 73 protei
8	132.5	10.1	517	13 R22904	1-Caldesmon. Gall
9	131.5	10.1	1898	17 Y30795	A human trichohyal
10	129.5	9.9	1132	17 R97866	Chicken leucocytos
11	126	9.6	1360	21 Y85263	Human protein kina
12	125	9.6	897	17 R92751	Murine EGF recepto

13	125	9.6	897	19 W47118	Murine eps15 prote
14	125	9.6	1297	20 W94406	Murine eps15 prote
15	124.5	9.5	1297	20 Y55932	Human ZC2 protein.
16	123	9.4	611	20 Y29039	T. gondii immunoge
17	122	9.3	1227	20 Y55965	Full length human
18	121.5	9.3	585	20 W97757	S-region transfer
19	121.5	9.3	1118	19 W82395	Human UBP protein
20	121	9.3	297	18 W20698	H. pylori cytoplas
21	121	9.3	412	17 W03626	Human thyrocytopin
22	120.5	9.2	542	18 W36490	Human TUBP1 protei
23	120.5	9.2	1299	21 Y58633	Protein regulating
24	120	9.2	360	17 W03627	Human follicle sti
25	120	9.2	432	20 W93954	Human regulatory m
26	119	9.1	185	20 W0129	Human endometrium
27	119	9.1	285	16 R82689	Shrimp tropomyosin
28	118	9.0	480	20 Y34653	Amino acid sequenc
29	118	9.0	1588	15 R46605	Malariat p16MP3 ep
30	118	9.0	1663	15 R46608	Plasmodium falciapa
31	116	8.9	411	20 W68011	Yeast immunophilin
32	116	8.9	424	20 Y49148	Amino acid sequenc
33	116	8.9	425	20 Y49147	Amino acid sequenc
34	116	8.9	426	20 Y49146	Amino acid sequenc
35	116	8.9	437	20 Y49241	N-terminal region
36	116	8.9	437	20 Y49244	N-terminal region
37	116	8.9	437	20 Y32180	N-terminal region
38	116	8.9	437	20 Y32183	N-terminal region
39	116	8.9	439	20 Y49240	N-terminal region
40	116	8.9	439	20 Y32179	N-terminal choline
41	116	8.9	530	19 W49042	Human low density
42	115.5	8.8	589	20 Y43392	PspC alpha-helix c
43	115.5	8.8	929	18 W45933	S. pneumoniae PspC
44	115.5	8.8	929	18 Y43384	S. pneumoniae PspC
45	115	8.8	223	20 W96760	A Helicobacter pyl

ALIGNMENTS

RESULT 1	
ID Y48610	standard; Protein: 168 AA.
XX Y48610:	
AC	
XX	
DT	08-DEC-1999 (first entry)
XX	
DE	Human breast tumour-associated protein 71.
XX	
KW	Expressed sequence tag; EST; human; breast; cancer; gene therapy;
KW	treatment; tumour; cytostatic; medicament.
XX	
OS	Homo sapiens.
XX	
PN	DE19813839-A1.
XX	
PD	23-SEP-1999.
XX	
PF	20-MAR-1998; 98DE-1013839.
XX	
PR	20-MAR-1998; 98DE-1013839.
XX	
PA	(META-) METAGEN GES GENOMFORSCHUNG MBH.
XX	
PI	Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E, Rosenthal A;
XX	WPI: 1999-528981/45.
DR	N-PSDB: 233669.
XX	
PT	Human nucleic acid sequences and protein products from tumor breast
PT	tissue, useful for breast cancer therapy -
XX	
PS	Claim 22; 174; 188pp; German.
XX	

FT	Modified-site	1012	/note= "potential phosphorylation site"
FT	Modified-site	1067	/note= "potential glycosylation site"
FT	Modified-site	1113	/note= "potential phosphorylation site"
PN	W0200006728-A2.		
XX			
PD	10-FEB-2000.		
XX			
PF	28-JUL-1999;	99WO-US17132.	
XX			
PR	28-JUL-1998;	98US-0123494.	
PR	14-SEP-1998;	98US-0152814.	
PR	14-OCT-1998;	98US-0173482.	
PR	03-NOV-1998;	98US-0106889.	
PR	19-NOV-1998;	98US-0109093.	
PR	22-DEC-1998;	98US-0113796.	
PR	12-JAN-1999;	98US-0173482.	
PR	12-JAN-1999;	99US-0229005.	
XX			
PA	(INCY-) INCYTE PHARM INC.		
XX			
P1	Hillman JL, Lal P, Tang YT, Corley NC, Guegler KU, Baughn MR;		
P1	Patterson C, Bandman O, Au-Young J, Gorgone GA, Yuen H, Azimzal Y;		
P1	Reddy R, Lu DAM, Shih LT;		
XX	WPI; 2000-183125/16.		
DR	N-PSDB; Z46153.		
XX			
PT	New human phosphorylation effectors useful for the diagnosis, treatment		
PT	and prevention of proliferative, immune and neuronal disorders -		
XX			
PS	Claim 1; Page 98-100; 142pp; English.		
XX			
CC	Y68769-95 and Y68797-99 represent human phosphorylation effectors (PHSP),		
CC	designated PHSP1-PHSP31 (The protein sequence for PHSP28 is not given		
CC	in the specification). The sequences were isolated from cDNA libraries		
CC	prepared from various human tissues. The PHSP proteins are useful for		
CC	the diagnosis, treatment and prevention of proliferative disorders,		
CC	immune disorders and neuronal disorders. The PHSP proteins form		
CC	pharmaceutical compositions which useful for treating or preventing		
CC	disorders associated with decreased PHSP expression/activity. PHSP		
CC	antagonists are useful for treating or preventing disorders associated		
CC	with increased PHSP expression/activity.		
XX			
SQ	Sequence 1135 AA:		
	Query Match	10.9%; Score 143; DB 21; Length 1135;	
	Best Local Similarity	19.0%; Pred. NO. 0.00068;	
	Matches 66; Conservative 41; Mismatches 83; Indels 156; Gaps		9
OY	21 LDERKKRQEQ---EMEKVRKPEDPECEPEEYDPRSL-----		54
	: : : : : : : : :		
Db	306 idrttkrkrgketeyysgeeeeeeypgeqepssivnpgestllrrdflrlqgenke		365
OY	55 -----YERLOEQRDKROQEYEDQFKFNMRGIDDEDTNLDVEYSROGLIENORRE		107
	: : : : : : : : : : :		
Db	366 rsealrrtqglldqegdlregeeykrlj-----laerqkrlieqkeqrirleeqqrter		417
OY	108 ELKELEKEYRNMLKKGISOENKKEYEKELTVKPIETKNFISOAKLLAGAVKHRSSESGNS		167
	: :		
Db	418 earlqge-----regrrrtegeekrllleelerrtrreeeertraaeekrrrvetagey		467
OY	168 VKRLKPD-----PEPDDKNOEPSSCKSLGNTLSGPSIHC		202
	: : : : : : : : : : : : : :		
Db	468 lrrtqlregegrnhlevlqqglldqegamllndhrtrpbhqhsqpppp-----qgersksfna		522
OY	203 P-----SAACIGIL-----		212
	: : : : : : :		
Db	533 pepkahyepadareavpvttsrpsvlstrdsplggsgqngsqdaqrnstslsleprrllwer		582

```

OY 213 -----GELGAYSGSDS-----ESSSDSFGT 233
                |||:||||:|
Db 563 veklvrpysgsssgsssgsqshpsqsgsggerfvrwsssksegs 630

RESULT 3
W30763
W30763 standard; Protein; 1178 AA.
AC W30763:
DT 07-MAY-1998 (first entry)
DE Mannose-1-phosphate transferase protein MNNA.
XX Mannose-1-phosphate transfer; MNNA gene; enzyme; yeast; regulatory gene;
XX human; high mannose type neutral saccharide chain.
XX Saccharomyces cerevisiae.
XX OS
XX JP09266792-A.
XX
XX 14-OCT-1997.
XX
XX 29-MAR-1996; 96JP-0075667.
XX
XX 29-MAR-1996; 96JP-0075667.
XX
XX (AGEN ) AGENCY OF IND SCI & TECHNOLOGY.
XX
XX WPI: 1997-553460/51.
XX N-PSDB: T91902.
XX
XX Positive regulatory gene of mannose-1-phosphate transfer in yeast -
XX useful for high mannose type neutral saccharide chain production
XX
XX Claim 1; Page 14-17; 23pp; Japanese.
XX
XX This sequence is encoded by the gene of the invention, designated MNNA,
XX and is a protein which positively regulates mannose-1-phosphate
XX transfer in yeast. The gene is useful for the preparation of human high
XX mannose type neutral saccharide chain.
XX
XX Sequence 1178 AA;

Query Match 10.9%; Score 142.5; DB 18; Length 1178;
Best Local Similarity 27.3%; Pred. No. 0.00079;
Matches 51; Conservative 38; Mismatches 61; Indels 37; Gaps 6;

OY 6 DGNLIKKRFVSEALDERRRKRRQ-----EEMKVKKRPDEPECPEVYVPRSLYERLQEO 61
      |||:||||:|
Db 1018 dkdpilvyedyayaklleerkrrekkkeeeekkkkeeeekkkke-----eekkkke 1070
      |||:||||:|
OY 62 KDRKOEEEOQKFFNMVRGLDEDTNPLDEYSROOELERORREBELKELKEYNRNLTK 121
      |||:||||:|
Db 1071 eeeekkkkeeeekkkke-----eekkkkeeeekkkkeeeekkkkeeeekkkke 1117
      |||:||||:|
OY 122 VGISQENKEVEKKLTVPETIKNFESQAKLLAGAVKHKSSSGNSVYKRLKDPPEPDKN 181
      |||:||||:|
Db 1118 e--deenkkkedeeekkkkeeeekkkgee-----knkkn-----deekkkgeeeekkn 1164
      |||:||||:|
OY 182 QEPSSCK 188
      |||:||||:|
Db 1165 eeeekkk 1171
      |||:||||:|

RESULT 4
R70491
R70491 standard; Protein; 562 AA.
CX
CX R70491;

```



```

Db 425 reqrrrgeekrrleerlrrekeerrrraeeekrrrereqeyrrlrqleeeqrhlvlg 484
QY 148 -----SQAKLAGAVKHKSSSGNSVKKRLKP-----PEPDDKNGEPSS 186
Db 485 qlldqegamlr--ecrvrmeehqrhaerlqrldqegayllslqdhrrprrpqsqqpppp 542
QY 187 CKSLGNTSLSGPSIHCP 203
Db 543 -----qgerskpsrthap 554

RESULT 6
Y55954
ID Y55954 standard: Protein; 1233 AA.
AC Y55954;
DT 18-FEB-2000 (first entry)
XX
DE Mouse STE20-related protein kinase NIK_m.
XX
KW Antithrombotic; antithrombotic; antithrombotic; antithrombotic; antithrombotic;
KW antiproliferative; antiproliferative; antiproliferative; antiproliferative; antiproliferative;
KW neuroprotective; cardiant; cerebroprotective; cytoprotective; antidiabetic;
KW vulnary; STE20; protein kinase; STK2; STK3; STK4; STK5; STK6; STK7;
KW ZC1, ZC2, ZC3, ZC4, KHS2, SUU1, SUU3, GSK2, PAK4, PAK5; antagonist;
KW antibody; gene therapy; rheumatoid arthritis; osteoarthritis; psoriasis;
KW inflammatory bowel disease; Crohn's disease; osteoarthritis; psoriasis;
KW rhinitis; autoimmunity; organ transplantation; multiple sclerosis;
KW myocardial infarction; cardiovascular disease; stroke; renal failure;
KW oxidative stress-related neurodegenerative disorder; Parkinson's disease;
KW amyotrophic lateral sclerosis; Leigh syndrome; cancer; cardiomyopathy;
KW ischemic disorder; inflammation; diabetes mellitus; fibrosis; mitosis;
KW mesangial disorder; growth regulation; wound healing; T cell activation;
KW immunosuppressant.
XX
OS Mus sp.
XX
PN WO953036-A2.
XX
PD 21-OCT-1999.
XX
PF 13-APR-1999; 99WO-US08150.
XX
PR 14-APR-1998; 98US-0081784.
XX
PA (SUGEN-) SUGEN INC.
XX
PI Plowman G, Martinez R, Whyte D;
XX
DR WPI; 1999-611301/52.
XX
PT Novel kinase-related polypeptides used for the diagnosis and treatment
XX of kinase-related diseases and disorders -
XX
PS Disclosure; Page 339-343; 387pp; English.
XX
XX
CC This sequence represents a novel STE20-related protein kinase. The
CC invention relates to nucleic acid molecule encoding a kinase polypeptide
CC selected from STK2, STK3, STK4, STK5, STK6, STK7, ZC1, ZC2, ZC3,
CC ZC4, KHS2, SUU1, SUU3, GSK2, PAK4 and PAK5. The proteins are used to
CC identify agonists and antagonists, and to raise antibodies. The
CC polynucleotides are useful in gene therapy protocols. The polynucleotides,
CC polypeptides, antibodies, antagonists and agonists may be used to treat
CC diseases such as immune-related disorders and diseases (e.g. rheumatoid
CC arthritis, atherosclerosis, chronic inflammatory bowel disease (e.g.
CC Crohn's disease), asthma, osteoarthritis, psoriasis, atherosclerosis,
CC rhinitis, autoimmunity, and organ transplantation, chronic inflammatory
CC peptic disease, multiple sclerosis, organ transplantation, myocardial
CC infarction, cardiovascular disease, stroke, renal failure, oxidative
CC stress-related neurodegenerative disorders (e.g. amyotrophic lateral
CC sclerosis, Parkinson's disease and Leigh syndrome), cancer,

```

```

CC cardiomyopathies, ischemic disorders, inflammatory disorders, diabetes
CC mellitus, fibrotic and mesangial disorders. The proteins may also be
CC useful for cell growth regulation (e.g. in wound healing), T cell
CC activation, mitosis control, and as immunosuppressants.
XX
SQ Sequence 1233 AA;
QY
Query Match 10.6%; Score 139; DB 20; Length 1233;
Best Local Similarity 23.6%; Pred. No. 0.0017;
Matches 55; Conservative 37; Mismatches 75; Indels 66; Gaps 8;
QY 21 LDERRRKROE---EWKVRKRPEDPECPPEVYDPRL----- 54
Db 305 idtrkrkgkdeleyeseeeevpegepsivvvpescltrflrlqgenke 364
QY 55 -----YERLQROKORQOEYEEQF-----KKRNVKGLDEDETFLDEVSROE 98
Db 365 rsealrrqqlldqegqlregeykrqlaerqkrleqkqegrrlrleeqqr-erearrqge 423
QY 99 LIEKORKEE---LKEIKETRNLIKVGISQENKKEVEKLTIV--KPIETRNK---FSQA 150
Db 424 reqrrrgeekrrleerlrrekeerrrraeeekrrrereqeyrrlrqleeeqrhlvlg 483
QY 151 KLAGAVKHKSSSGNSVKKRLKPDPEDDKNGEPSSCKSLGNTSLSGPSIHCP 203
Db 484 qlldqegamlhrrhrprrpqsqqppppqdr-----skpsrthap 521

RESULT 7
Y58500
ID Y58500 standard: Protein; 1162 AA.
AC Y58500;
DT 10-APR-2000 (first entry)
XX
DE HHV8 ORF 73 protein, SEQ ID NO:21.
XX
KW HHV8; detection; diagnosis; Kaposi's sarcoma; AIDS; immunogen;
KW antigen.
XX
OS Human herpesvirus type 8.
XX
FH Key Location/Qualifiers
FT Misc-difference 96
FT /Label= unknown
XX
PN WO961909-A2.
XX
PD 02-DEC-1999.
XX
PF 26-MAY-1999; 99WO-US11407.
XX
PR 26-MAY-1998; 98US-0086695.
XX
PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI Pau CP;
XX
DR WPI; 2000-097142/08.
XX
PT New methods and compositions for the detection of human herpesvirus -
XX Claim 2; Page 59-62; 68pp; English.
XX
XX
CC Sequences Y58480-Y58532 represent immunogenic polypeptides derived from
CC human herpes virus type 8 (HHV8, a gammaherpesvirus). HHV8 plays an
CC important role in the pathogenesis of AIDS-related Kaposi's sarcoma. The
CC invention relates to a novel method of detecting the presence of human
CC herpesvirus 8 in a biological sample using peptides representative of
CC dominant antigenic regions of HHV8. The method comprises contacting one
CC or more isolated, immunogenic HHV8 peptides with an antibody-containing

```

CC biological sample, and detecting the formation of a complex between the
CC peptide and the antibody. The presence of a peptide-antibody complex
CC indicates the presence of human herpesvirus 8. The detection of HHV8
CC infection can be used to diagnose AIDS-associated Kaposi's sarcoma. The
CC HHV8-specific antibodies are useful therapeutically when for the passive
CC immunisation of a human against HHV8 infection, thereby reducing HHV8
CC related disease. The detection and treatment of Kaposi's sarcoma could
CC accurate. Early detection and treatment of Kaposi's sarcoma could
CC diminish the severity of symptoms related to AIDS and the sensitive
CC techniques could reduce erroneous characterisations of skin disorders.
CC Previous assays for HHV8 antibodies such as immunofluorescence assays,
CC immunoblots and enzyme immunoassays lack the sensitivity and accuracy
CC needed for reliable diagnosis of Kaposi's sarcoma. Further advantages
CC of the assays are that reproducible results are obtained and the method
CC is suitable for rapid throughput and screening of samples economically.
CC
XX
SQ Sequence 1162 AA;

Query Match 10.2%; Score 133; DB 21; Length 1162;
Best Local Similarity 25.5%; Pred. No. 0.0051;
Matches 41; Conservative 41; Mismatches 61; Indels 18; Gaps 4;

OY 18 EAELDERKKROEEKVKRPDPPECEPEVYDPRSLYERLOEQDRKOQTEEOFKRN 77
| | | | | : : : : : | | | | : : : : : | |
Db 836 egeleegeveegeveegeveegevee-----eegeveegeveegevee----- 883
OY 78 MVRGLEDDETFLNDEVSRQOELIEKORREELKEKEYRNNNKKGISQENKKEVEKKLT 137
| | | | | : : : : : | | | | : : : : : | |
Db 884 -----veegeveegeveegeveegeveegeveegeveegeveegeveepl 938
OY 138 VKPIETKPKFS-QAKTLGAVKHKSSSGSNVYKRLKPPPEP 177
| | | | | : : : : : | | | | : : : : : | |
Db 939 lhgsssedemevdyprvstheqlassppgdntppddppgp 979

RESULT 8
R22904 R22904 standard; Protein; 517 AA.

XX AC R22904;

XX DT 30-JUL-1992 (first entry)

XX DE 1-Caldesmon.

XX KM Smooth muscle; contraction.

XX OS Gallus gallus.

XX PN JP03240798-A.

XX PD 28-OCT-1991.

XX PF 20-FEB-1990; 90JP-0037362.

XX PR 20-FEB-1990; 90JP-0037362.

XX PA (TAKA-) TAKARA SHUZO KK.

XX DR WPI; 1992-107714/14.

XX DR N-PSDB; Q23748.

XX XX

PT 1-Caldesmon polypeptide - used for the regulation of smooth muscle
contraction

XX PS Claim 2; Fig 2; 7pp; Japanese.

CC The protein sequence was deduced from the DNA sequence obtd. by
screening a chicken embryo cDNA library. The protein encoded by
CC recombinant 1-caldesmon is involved in the regulation of
CC smooth muscle contraction.

XX

SQ Sequence 517 AA;

Query Match 10.1%; Score 132.5; DB 13; Length 517;
Best Local Similarity 23.5%; Pred. No. 0.0019;
Matches 60; Conservative 40; Mismatches 86; Indels 69; Gaps 11;

OY 18 EAELDERKKROEEKVKRPDP-----PECEPEVYDPRSLYER 57
| | | | | : : : : : | | | | : : : : : | |
Db 28 eaarerrrrrrr-----erlrqkeegvsgevtksevnqnsvaeectrrstddeall 84
OY 58 LOEQDRKOQTEYEF-----QKFKNNV-----RGLDEDTFLNDEVSRQOELIEKO 103
| | | | | : : : : : | | | | : : : : : | |
Db 85 larrreerrgkrlqealergkefdpftldgslyvpreravnneen-----eltgkeekve 141
OY 104 RREELAKE-----LKEY-RNNLKKGISQENKKEVEKKLTVPKRNKPSOAKLLAGAVK 158
| | | | | : : : : : | | | | : : : : : | |
Db 142 gqrceleeetvtskyqrnnwrgdg-----eeegkkekdegeqekpvtceagvk----- 193
OY 159 HKSSSGSNVYKRLKPPPEP-----DDKNOEPSSCKSLGNTSLGSPSHGPSAAVCIGILPG 214
| | | | | : : : : : | | | | : : : : : | |
Db 194 -----dhkvkekpkkeemkswdtkrgvpeqkagngertltpkvtetnaf-----g 241
OY 215 LGAYSGSSDSESSSD 229
| | | | | : : : : : | | | | : : : : : | |
Db 242 rsnlkgaaanaeagse 256

RESULT 9

Y30795 Y30795 standard; Protein; 1898 AA.

XX AC Y30795;

XX DT 25-NOV-1999 (first entry)

XX DE A human trichohyalin (TRHY) protein.

XX KM Human; trichohyalin; TRHY; protein; tissue structure; wound healing;
terminally differentiating epidermal tissue; proteinaceous gel;

XX KW breast implant.

XX OS Homo sapiens.

XX PN US5958752-A.

XX PD 28-SEP-1999.

XX PF 14-FEB-1997; 97US-0800644.

XX PR 30-APR-1993; 93US-0056200.

XX PA (USSH) US DEPT HEALTH & HUMAN SERVICES.

XX PI Kim I, Chung S, Park S, Steinert PM, Lee S;

XX DR WPI; 1999-561041/47.

XX DR N-PSDB; 222301.

XX PT Human trichohyalin useful for forming a proteinaceous gel that promotes
wound healing -

XX PS Disclosure; Fig 3A-W; 126pp; English.

CC The present sequence represents a human trichohyalin (TRHY) protein.
The protein is found in terminally differentiating epidermal tissue,
and is involved in forming the structural architecture of such
tissue. The trichohyalin protein is useful for forming a
proteinaceous gel which may then be used for healing wounds, or in
breast implants.

XX CC

XX SQ Sequence 1898 AA;

```

OY 124 ISOENK-----KEVEKKLFVFKPIETNNKFQOALLGAVNKKSESGNSVRLK 172
      :| | :| | :| | :| | :| | :| | :| | :| | :| | :| | :| | :| |
Db 1011 heeekvhtheeenvtlyeeekvhtheeek-vmkkkkimykkkkkkkkggag--- 1066
OY 173 PDPEPDKNQEP--SSCKSLGNTSLSGPSIHCPs 204
      | | | | | :| | :| | :| | :| | :| | :| | :| | :| | :| |
Db 1067 ---vpgsnspblestcrg-----awhps 1088

RESULT 11
ID Y85263 standard; Protein; 1360 AA.
AC Y85263:
XX
XX 29-JUN-2000 (first entry)
DE Human protein kinase KIAA0551 amino acid sequence.
KW Protein kinase; KIAA0551; neuropathy; neuropathic pain; inflammation;
KM chronic pain; neurodegenerative disorder; neurotraumatic disorder;
XX Parkinson's disease; Alzheimer's disease; ischemic disease.
XX Homo sapiens.
XX
XX Key Location/Qualifiers
FH Misc-difference 290 /note= "Encoded by GA"
FT FT
XX MOZ00015805-A1.
XX PD 23-MAR-2000.
XX PF 10-SEP-1999; 99WO-GB03017.
XX PR 10-SEP-1998; 98GB-0019779.
XX PR 29-MAR-1999; 99GB-0007261.
XX PA (SMIK ) SMITHKLIN BEECHAM PLC.
X1 Bingham S., Case P., Lawson SN, Newton RA, Rausch OL, Relth AD;
P1 Sanger GJ;
XX WP1: 2000-271443/23.
DR N-PSDB: A10669.
PT Isolated human KIAA0551 polynucleotide and polypeptide, useful for
PT treating e.g. neuropathies, neuropathic pain, inflammatory and chronic
PT pain and neurodegenerative conditions -
PS Claim 2; Page 41; 48pp; English.
XX
XX This sequence represents the human protein kinase KIAA0551 amino acid
CC sequence. The nucleotide sequence was isolated from a human foetal brain
CC cDNA library. The rat KIAA0551 mRNA is upregulated in dorsal root ganglia
CC during sciatic neuropathy (a procedure accompanied by increased
CC sensitivity to somatic pain) which indicates a role for KIAA0551 in the
CC regulation of molecular processes associated with neuropathy and
CC neuropathic pain. The KIAA0551 protein may be used for treating
CC neuropathies, neuropathic pain, inflammatory and chronic pain,
CC neurodegenerative conditions such as Parkinson's disease and Alzheimer's
CC disease, and neurotraumatic disease or ischemic disease damage in
CC cardiac tissue.
XX
XX Sequence 1360 AA:

Query Match 9.6%; Score 126; DB 21; Length 1360;
Best Local Similarity 23.1%; Pred. No. 0.025;
Matches 54; Conservative 50; Mismatches 68; Indels 62; Gaps 10.

21 LDERKKRROEWEKVRKPEDPECEPEEV-----YDPRSLYERLQEKRKQOEYEOPK 74

```


CC pelvic disease, multiple sclerosis, organ transplantation, myocardial
CC infarction, cardiovascular disease, stroke, renal failure, oxidative
CC stress-related neurodegenerative disorders (e.g. amyotrophic lateral
CC sclerosis, Parkinson's disease and Leigh syndrome), cancer,
CC cardiomyopathies, ischemic disorders, inflammatory disorders, diabetes
CC mellitus, fibrotic and mesangial disorders. The proteins may also be
CC useful for cell growth regulation (e.g. in wound healing), T cell
CC activation, mitosis control, and as immunosuppressants.

XX
SQ Sequence 1297 AA;

Query Match 9.5%; Score 124.5; DB 20; Length 1297;
Best Local Similarity 25.0%; Pred. No. 0.032;
Matches 54; Conservative 43; Mismatches 64; Indels 55; Gaps 10;

OY 21 LDERKKRROEWEKRPEDPECEPEVYDPRSLYERLQEKDRKO-QEYEQFVKKNV 79
DB 354 laerqkrlleegkqrrlleegqrreke-----lrqgreqrthyeeqmrreer 403
OY 80 RGLDEDTNFDLDEVSROELIEKORREEL-KELKEYRNMLKKVGISQENKKEVEKK--- 135
DB 404 rraehq-----eykrkg--leeqrqaerlqrqlkqerdyivsldqrdqepvkekply 456
OY 136 -----LTVKPIETKNKFSQAKLLAGAVKHK-----SSESGNS 167
DB 457 hykegmspsekpwakeveerslrrgs--spamphkvanrisdpnlprsesfsisgyq 514
QY 168 VKRLKPDPDPDKNOEPS--SKSLGNTSLSGPSIH 201
DB 515 partppmlrpydp-qiphlvavksgqpaltasgvh 549

Search completed: February 16, 2001, 17:05:33
Job time: 27 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2001, 17:05:07 ; Search time 15.5 Seconds
(without alignments)
1112.693 Million cell updates/sec

Title: US-09-602-597-2

Perfect score: 1306

Sequence: 1 MDGGDDGLIKRRVSEAE.....NATGKIVSIFRTNPLEAP 254

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 195891 seqs, 67900655 residues

Total number of hits satisfying chosen parameters: 195891

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

PIR66:*
1: PIR1:*
2: PIR2:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	302.5	23.2	219	2 T19438	hypothetical prote
2	169.5	13.0	241	2 T48006	hypothetical prote
3	158	12.1	1240	2 S52734	hypothetical prote
4	153	11.7	707	2 S60588	hypothetical prote
5	151	11.6	649	2 JN0809	drebrin A - rat
6	151	11.6	651	2 T14763	drebrin E (clone g
7	149.5	11.4	312	2 T25994	hypothetical prote
8	149	11.4	1087	2 T30330	hypothetical prote
9	146	11.2	292	2 JEO233	gelsoin-related p
10	146	11.2	451	2 G70241	tropomyosin-I - scal
11	142.5	10.9	1178	2 S76475	hypothetical prote
12	142	10.9	772	2 T50463	mannosylphosphoryl
13	142	10.9	1407	1 S28589	protein kinase - c
14	141.5	10.8	466	2 T22141	trichohyalin - rab
15	141	10.8	1027	2 T46481	hypothetical prote
16	139.5	10.7	564	2 B43776	hypothetical prote
17	139	10.6	1233	2 T30989	drebrin E1 - chick
18	138.5	10.6	335	2 T33457	serine/threonine p
19	138	10.6	385	2 T19201	hypothetical prote
20	137	10.5	1359	2 T34036	hypothetical prote
21	137	10.5	678	2 A54514	glutamic acid-rich
22	136.5	10.5	729	2 T50989	hypothetical prote
23	135	10.3	777	2 T21048	hypothetical prote
24	135	10.3	1526	2 A45605	hypothetical prote
25	133	10.2	1526	2 A45605	mature-parasite-in
26	132.5	10.1	839	2 T137356	epithelial microtu
27	132.5	10.1	839	2 T137356	class II INCENP pro
28	132.5	10.1	1737	2 A59235	unconventional myo
29	132	10.1	405	2 T16922	hypothetical prote

30	132	10.1	2058	2 A59267	myosin X - human
31	131.5	10.1	517	1 A39038	1-caldesmon, nonmu
32	131.5	10.1	1898	1 A45973	trichohyalin - hum
33	131	10.0	791	2 T24435	hypothetical prote
34	131	10.0	1038	2 T02634	rep protein homolo
35	130.5	10.0	1403	2 T11583	probable translati
36	130	10.0	241	2 S44893	ZK1236.7 protein -
37	130	10.0	550	2 A46419	trophoblast-endoth
38	130	10.0	695	2 T54325	gene XE7 protein -
39	129.5	9.9	665	2 B71609	hypothetical prote
40	129.5	9.9	2052	2 T18519	myosin X - bovine
41	129	9.9	812	2 S43604	R0755.1 protein (c
42	128	9.8	805	2 E70474	translation initia
43	128	9.8	1233	2 T54383	chromosome segrega
44	128	9.8	1997	2 T71607	DNA helicase II BR
45	127.5	9.8	301	2 T33068	hypothetical prote

ALIGNMENTS

```
RESULT 1
T19438
hypothetical protein C25A1.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T19438
R:Northmore, B.
submitted to the EMBL Data Library, October 1996
A:Reference number: Z19124
A:Accession: T19438
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-219 <WILD>
A:Cross-references: EMBL:Z81038; PIDN:CAB02762.1; GSPDB:GN00019; CESP:C25A1.1
A:Experimental source: clone C25A1
C:Genetics:
A:Gene: CESP:C25A1.1
A:Map position: 1
A:Introns: 33/1; 66/1; 151/2

Query Match 23 2%; Score 302.5; DB 2; Length 219;
Best Local Similarity 36.4%; Pred. No. 816-11;
Matches 82; Conservative 36; Mismatches 70; Indels 37; Gaps 8;

QY 15 FVSEAELEDERKRRROEMEKVRKPEDEECPEEYVPSLYERLOQDKRQOEYEOFK 74
    |||:||||:| ||| |||:|||| | | | :||:|: | | | | | | |
Db 5 FVSTSELEDEKRRQOEMEKIRKPTDATTVPPEYCNKTLFEDLKNKKDAKOLEIDAKK 64

QY 75 FKNMVGLEDETFNPLDEVSROQELIEKORREBELKELKYRNLLKKGVISQE--NKKEV 132
    |||||:||||:| | | :|:|:| | | | | | | | | | | | |
Db 65 LKNMVGIDEDSVFLSELDSTKRVA-KMKMKREDEQEL-----IKELAVTQHLAANQS 117

QY 133 EKKLYTKPIETK-----NKPSQAKLLAGAVYKKSSEGSNYSKRLKPPPEPDKNQOESSCK 188
    |||||:||||:| | | :|:|:| | | | | | | | | | | | |
Db 118 SSRFTLKPTSKVLGPCKSKQAALFSTAIKRTST-----EKKRKE----- 159

QY 189 SLGNTSLSGPSIHCPAAVCGITLPGICAVSGSDSES--SSDSE 231
    : : : | | | | | | | | | | | | | | | | | | |
Db 160 DVVSSKSVKRE----FYIKQIGALQALCEYDPSDSESDASSDSE 200

RESULT 2
T48006
hypothetical protein T17J13.100 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000
C:Accession: T48006
R:Rieger, M.; Mueller-Auer, S.; Zipp, M.; Schaefer, M.; Mewes, H.W.; Lemke, K.; Maye
submitted to the Protein Sequence Database, February 2000
A:Reference number: Z24482
A:Accession: T48006
```


C:Comment: This protein is an actin-binding protein.
C:Genetics:
A:Map position: 5
C:Keywords: actin binding

Query Match 11.6%; Score 151; DB 2; Length 649;
Best Local Similarity 25.8%; Pred. No. 0.11;
Matches 49; Conservative 44; Mismatches 61; Indels 36; Gaps 7;

QY 17 SEAELEDERRRKROEWEKVRKPPDEPECEPEYVDPRLSLYERLOEQDKRQOEYEDQFKR 76
DB 166 TDAVEMKRRINRQEFWEQAK--EELRKEERKRALDERLRFEOERMEOEKQEO---- 218
QY 77 NMVRLGDEDETNFLDEVSRQOELIEKORREBELKELEYRNNLKVGISOENKKEVEKKL 136
DB 219 -----EEERRRIR--REQQ-TEHRRKQOTLEAEAKRRKLEQOSTIGDHDDEE--- 265
QY 137 TVKPIETKKNKFSQAKLLAGAVKHKSSGNSVYRLKPD-PEPDDKNQEPSSCKSLGNTSL 195
DB 266 -----ETHMKKSESEV-----EFAAATIAQRPDNPREFFKQOEYASASAGSCDV 310
QY 196 GSPSIHCPSA 205
DB 311 PSPFNHRPGS 320

RESULT 6

hypothetical protein DKFp434D064.1 - human
C:Species: Homo sapiens (man)
C>Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
R:Wambutt, R.; Heubner, D.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, August 1999
A:Reference number: Z18181
A:Accession: T14763
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-651 <MAN>
A:Cross-references: EMBL:AL110225
A:Experimental source: adult testis; clone DKFp434D064
C:Genetics:
A>Note: DKFp434D064.1

Query Match 11.6%; Score 151; DB 2; Length 651;
Best Local Similarity 25.8%; Pred. No. 0.11;
Matches 49; Conservative 44; Mismatches 61; Indels 36; Gaps 7;

QY 17 SEAELEDERRRKROEWEKVRKPPDEPECEPEYVDPRLSLYERLOEQDKRQOEYEDQFKR 76
DB 168 TDAVEMKRRINRQEFWEQAK--EELRKEERKRALDERLRFEOERMEOEKQEO---- 220
QY 77 NMVRLGDEDETNFLDEVSRQOELIEKORREBELKELEYRNNLKVGISOENKKEVEKKL 136
DB 221 -----EEERRRIR--REQQ-TEHRRKQOTLEAEAKRRKLEQOSTIGDHDDEE--- 267
QY 137 TVKPIETKKNKFSQAKLLAGAVKHKSSGNSVYRLKPD-PEPDDKNQEPSSCKSLGNTSL 195
DB 268 -----ETHMKKSESEV-----EFAAATIAQRPDNPREFFKQOEYASASAGSCDV 312
QY 196 GSPSIHCPSA 205
DB 313 PSPFNHRPGS 322

RESULT 7

hypothetical protein ZK354.3 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T25994

R:Johnson, D.; Wamsley, P.; Bradshaw, H.
submitted to the EMBL Data Library, February 1997
A:Description: The sequence of C. elegans cosmid ZK354.
A:Reference number: Z20120
A:Accession: T25994
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-312 <JOB>
A:Cross-references: EMBL:U88172; PIDN:AAB42258.1; GSPDB:GN00022; CESP:ZK354.3
A:Experimental source: strain Bristol N2; clone ZK354
C:Genetics:
A:Gene: CESP:ZK354.3
A:Map position: 4
A:Introns: 41/1; 226/3

Query Match 11.4%; Score 149.5; DB 2; Length 312;
Best Local Similarity 24.9%; Pred. No. 0.061;
Matches 46; Conservative 42; Mismatches 78; Indels 19; Gaps 5;

QY 20 ELDERKRRKROEWEKVRKPPDEPECEPEYVDPRLS-----LYERLOEQDKRQ 66
DB 68 EKKEEKKEEKKKEEKKKADDEKKTKEKDKKSKKTEKDKLSVKKTOETSERDKKD 127
QY 67 OEYEQFKFKNMVRGLDEDETNFLDEV--SRQOELIEKORREBELKELEYRNNLKVG 124
DB 128 ERKEDDKKEENKESKDEKKK--DEVKDKDEKMDKKPGKEKKKEEFKEMKKEK 185
QY 125 SQENKKEVEKK-LTVKPIETKKNKFSQAKLLAGAVKHKSSGNSVYRLKPDPEDDKNOE 183
DB 186 KEKKKEPKKNDAPKKEGFKGVKDALVENPIYEMSDR-DEKKEKKDKDKDEKDE 244
QY 184 PSSCK 188
DB 245 KREAK 249

RESULT 8

gelsoilin-related protein GRP125 - slime mold (Dictyostelium discoideum)
C:Species: Dictyostelium discoideum
C>Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 22-Oct-1999
C:Accession: T30330
R:Stocker, S.; Hiery, M.; Marriot, G.
Mol. Biol. Cell 10; 161-178, 1999
A:Title: Phototactic migration of Dictyostelium cells is linked to a new type of gels
A:Reference number: Z20823; MUID:99096692
A:Accession: T30330
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1087 <STO>
A:Cross-references: EMBL:U95159; NID:94100185; PID:94100186; PIDN:AAD0774.1
C:Genetics:
A:Introns: 137/1

Query Match 11.4%; Score 149; DB 2; Length 1087;
Best Local Similarity 24.1%; Pred. No. 0.25;
Matches 59; Conservative 59; Mismatches 95; Indels 32; Gaps 9;

QY 11 IKRRFVSEAELEDERRRKROEWEKVRKPPD--PEPCPEYVDPRLSLYERLOEQDKRQ 67
DB 717 LAERLOKEKEDLEKLOEQOQOQKNNKIVEEVEVKE-EDVKEEVEEVEEKEE 775
QY 68 EYEQFK-----FKNMVRGLDEDETNFLDEVSRQOELIEKORREBELK-ELK--EYRNN 118
DB 776 VKEEYKVEAKETKEIEEYVNDATYEVKNQVVEEVEEVEEVEEVEEVEEVEEVEE 835
QY 119 LKKVGISOENKKEVEKKLTVKPI--ETKKNFSQAK--LLAGAVKHKSSGNSVYRLKPD 174
DB 836 AKEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEE 895
QY 175 PEPDDKNQEPSSCKSLGNTSLGSPSIHCPSAVVCIGILPLGAYSSSSSSSSSDESGT 234

Db 896 EKVEENETVNEVEVG-IIVSPSEKVEA-----NSSSTISSPENEGSV 940
 QY 235 NATGK 239
 Db 941 SVKDK 945

RESULT 9

JE0233
 troponin-I - scallop (*Chlamys nipponensis*)
 C:Species: *Chlamys nipponensis* (Japanese scallop)
 C:Date: 05-Dec-1998 #sequence_revision 05-Dec-1998 #text_change 17-Mar-2000
 C:Accession: JE0233
 R:Tanaka, H.; Ojima, T.; Nishita, K.
 J. Biochem. 124, 304-310, 1998
 A:Title: Amino acid sequence of troponin-I from Akazara scallop striated adductor muscle
 A:Reference number: JE0233; MUID:98351986
 A:Accession: JE0233
 A:Molecule type: protein
 A:Residues: 1-292 <TAB>
 C:Superfamily: troponin T

Query Match 11.2%; Score 146; DB 2; Length 292;
 Best Local Similarity 28.1%; Pred. No. 0.09;
 Matches 47; Conservative 26; Mismatches 62; Indels 32; Gaps 4;

QY 18 EAELDERKKRQOEEMEKRRPEDECEVEYDPRSLYERLOEQDKRQOEYEDFKFN 77
 Db 55 DAELNRRRQOEELAAARAE-----YNRQOEELRQOEERQRRREE 102
 QY 78 MVRGLDEDTNPLDEVSRQOELEKORREELKELKEYRNMLKV--GISOENKKEYEK 134
 Db 103 QRQOEERERLRER-----EQEREERARRAEQKKKKKGGLGSPKKMLLK 154
 QY 135 KLYKPIETKPKFSQAKLLAGAVKHKSSGNSVYKRLKPDPEPDDKN 181
 Db 155 LIMOKAED-----LANEAKAKAEKKEKYINDLVKPFSTGKD 192

RESULT 10

G70241
 hypothetical protein Bb16 - Lyme disease spirochete plasmid I/1p28-4
 C:Species: *Borrelia burgdorferi* (Lyme disease spirochete)
 C:Date: 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 08-Oct-1999
 C:Accession: G70241
 R:Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; White
 son, D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vugt,
 ; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.
 Nature 390, 580-586, 1997
 A:Authors: Smith, H.O.; Venter, J.C.
 A:Title: Genomic sequence of a Lyme disease spirochete, *Borrelia burgdorferi*.
 A:Reference number: A70100; MUID:98065943
 A:Accession: G70241
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-451 <KLE>
 A:Cross-references: GB:AE000789; NID:g2650079; PID:MAC66203.1; PID:g2690100; TIGR:BB116
 A:Experimental source: strain B31
 C:Genetics:
 A:Genome: plasmid

Query Match 11.2%; Score 146; DB 2; Length 451;
 Best Local Similarity 25.1%; Pred. No. 0.14;
 Matches 46; Conservative 46; Mismatches 77; Indels 14; Gaps 4;

QY 13 KRFSAEALDERRRRQOEEMEKRRPEDECEVEYDPRSLYERLOEQDKRQOE--- 68
 Db 44 KQKSKPRTEELKKQOE-BELKKQOEELKKQOEELKKQOEELKKQOEELKK 102
 QY 69 --YEOFKFNWVGLEDDETNPFLDEVSRQOELEKORREELKELKEYRNMLKVGIS 125

Db 103 KROOEELKKQOEELKKQOEELKKQOEELKKQOEELKKQOEELKKQOEELKKQOE 161
 QY 126 QENK-----EVEKRLTVKPIETKPKFSQAKLLAGAVKHKSSGNSVYKRLKPDPEPDDK 180
 Db 162 EELKKQOEELKKQOEELKKQOEELKKQOEELKKQOEELKKQOEELKKQOEELKK 221

QY 181 NOE 183
 Db 222 QOE 224

RESULT 11

S78475
 mannosylphosphorylation protein MNNA - Yeast (*Saccharomyces cerevisiae*)
 N:Alternate names: protein YKL200c; protein YKL201c
 C:Species: *Saccharomyces cerevisiae*
 C:Date: 11-Mar-1998 #sequence_revision 17-Apr-1998 #text_change 05-Jan-1998
 C:Accession: S78475; S38037; S38038
 R:Odani, T.; Shima, Y.; Yoshitani, J.
 submitted to the EMBL data library, January 1996
 A:Description: Cloning and Analysis of the MNNA Gene Required for Phosphorylation of
 A:Reference number: S78475
 A:Accession: S78475
 A:Molecule type: DNA
 A:Residues: 1-1178 <ODA>
 A:Cross-references: EMBL:D81006; NID:g1752735; PID:d1012343; PID:g1752736; MIPS:YKL20
 R:Ma e Silva, A.; Bossier, P.; Vilela, C.; Fernandes, L.; Soares, H.; Guerreiro, P.
 submitted to the Protein Sequence Database, March 1994
 A:Reference number: S38024
 A:Accession: S38037
 A:Molecule type: DNA
 A:Residues: 121-249, 'A10LOT', 255, 'MALTRD', 262-390, 'LRISNSN', 398-515, 'LG' <MAI>
 A:Cross-references: EMBL:Z28200; NID:g486355; PID:g486356
 A:Experimental source: strain S288C
 A:Note: this sequence has been revised in reference S78475
 A:Note: this was assumed to be protein YKL200c
 A:Accession: S38038
 A:Molecule type: DNA
 A:Residues: 640-1178 <MAV>
 A:Cross-references: EMBL:Z28201; NID:g486357; PID:g486359; MIPS:YKL201c
 A:Experimental source: strain S288C
 A:Note: this sequence has been revised in reference S78475
 A:Note: this was assumed to be the complete sequence of protein YKL201c
 C:Genetics:
 A:Gene: MNNA
 A:Cross-references: MIPS:YKL201c
 A:Map position: 11L
 A:Function:
 A:Description: required for phosphorylation of N-linked oligosaccharides
 C:Keywords: transmembrane protein
 F:28/44/Domain: transmembrane #status predicted <TM>

Query Match 10.9%; Score 142.5; DB 2; Length 1178;
 Best Local Similarity 27.3%; Pred. No. 0.63;
 Matches 51; Conservative 38; Mismatches 61; Indels 37; Gaps 6;

QY 6 DGNLIKKRFSAEALDERRRRQOEEMEKRRPEDECEVEYDPRSLYERLOEQ 61
 Db 1018 DKDPIIYEDYAVAKLLEERKRREKKKEEKKKEE-----EKKRKE 1070
 QY 62 KDRQOEYEQDFKFNWVGLEDDETNPFLDEVSRQOELEKORREELKELKEYRNMLKK 1
 Db 1071 EEEKKKEEEKKKE-----EEKKQOEELKKQOEELKKQOEELKKQOEELKK 1117
 QY 122 VGISOENKKEYEKRLTVKPIETKPKFSQAKLLAGAVKHKSSGNSVYKRLKPDPEPDDKN 181
 Db 1118 E--DEENKKNDEKKKNEEKKKEOE-----KNKNE--DEKKKQOEELKK 1164
 QY 182 QEPSSCK 188

Db 1165 EEEERKK 1171

RESULT 12

150463 protein kinase - chicken

C.Species: Gallus gallus (chicken)

C.Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 24-Sep-1999

C.Accession: I50463

R.I.L., H.; Grenet, J.; Valentine, M.; Lahti, J.M.; Kidd, V.J.

Gene 153, 237-242, 1995

A.Title: Structure and expression of chicken protein kinase PITSRE-encoding genes.

A.Reference number: I50463; MUID:95180725

A.Accession: I50463

A.Status: preliminary; translated from GB/EMBL/DBJ

A.Molecule type: DNA

A.Residues: 1-772 <LTX>

A.Cross-References: EMBL:U16656; NID:g571458; PIDN:AAA67037.1; PID:g571460

C.Genetics:

A.Gene: cPITSRE

A.Introns: 37/3; 76/2; 118/1; 164/2; 208/1; 249/1; 283/1; 329/1; 351/1; 407/3; 438/1; 47

C.Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homolo

C.Keywords: ATP

F:413-668/Domain: protein kinase ATP-binding motif

F:421-429/Region: protein kinase ATP-binding motif

Query Match 10.9%; Score 142; DB 2; Length 772;

Best Local Similarity 23.9%; Pred. No. 0.43;

Matches 63; Conservative 50; Mismatches 79; Indels 72; Gaps 12;

```

QY 3 GGDGDLIIRK-----KRFVSEALDERRRKRRQ-----EMEKVKRPEDPECEVVD 50
Db 77 GEEDDSLAIKPPQOQKKEKTHHRRKDEKRRHSHSNAVAVKKEKRF-----HR 131
QY 51 PRSLYRLQOKRQOEYEQFKFNMYAGLDETFNLFDEVSROOELIEKORREE 108
Db 132 RR---KRHEEDQKARREMRQ-KREMARHSHRREDRLEQERREKIRKQOEQRE 187
QY 109 LKELKRYNNLKKVGIQENKKEVE-----KLLTVKPIETKKNFSQAKLLAGAV 157
Db 188 OKERERRAEERRK---EREARRVSAHHHTVREYGDVYMRP-----WSRSLRQORD 238
QY 158 KHKSSSGSVKRLKRP---DPEPD-----DKNOEPSSCKSLGTSISGPIHPSAAYCI 209
Db 239 KRPQASRKRPVKEEKPEERDPLSDLDISDEKRTSSAKMSLQASG----- 286
QY 210 GILPGLGAYSGSSSDSSDSEGT 233
Db 287 -----SGSEEEESSSESS 301

```

RESULT 13

S28589 trichohyalin - rabbit

C.Species: Oryctolagus cuniculus (domestic rabbit)

C.Date: 12-Mar-1993 #sequence_revision 01-Mar-1996 #text_change 22-Jun-1999

C.Accession: S28589

R.Fietz, M.J.; Rogers, G.E.

submitted to the EMBL Data Library, December 1992

A.Description: Examination of the gene encoding rabbit trichohyalin.

A.Reference number: S28589

A.Accession: S28589

A.Molecule type: DNA

A.Residues: 1-1407 <FIE>

A.Cross-References: EMBL:Z19092; NID:g1746; PIDN:CAA79519.1; PID:g1747

C.Comment: Trichohyalin is a protein of the medulla of the hair and of the inner root sh

C.Genetics:

A.Introns: 46/3

C.Superfamily: trichohyalin; calmodulin repeat homology

C.Keywords: calcium binding; citrulline; EF hand; hair; tandem repeat

F:49-81/Domain: calmodulin repeat homology <EF2>

Query Match 10.9%; Score 142; DB 1; Length 1407;

Best Local Similarity 26.9%; Pred. No. 0.81;

Matches 49; Conservative 39; Mismatches 48; Indels 46; Gaps 8;

```

QY 13 KRFVSEALDERRRKRR-----QEEMKY-----RKPEDECEPEYDPSRYERI 58
Db 215 KRELREE-QQRREREQERLQEEEDQLRORRRREPREQQLRLEELIREQRRL 273
QY 59 OEQDKRQO-----EYEB--QFKFNMYRGIDEFTNLFDEVSROOELIEKORREE 109
Db 274 EGERREQLRRQRLQERGERRQQLRLEELIREQRRLQERREQLRQERREQL 333
QY 110 K-ELKRYNNLKKVGIQENKKEVEKLLVPIETKKNFSQAKLLAGAVKHSSEGSNV 168
Db 334 KRELREIREQR--LEQERRE-----QLLAEEVQARERGESL 372
QY 169 KR 170
Db 373 TR 374

```

RESULT 14

T22141 hypothetical protein F43G9.10 - Caenorhabditis elegans

C.Species: Caenorhabditis elegans

C.Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

C.Accession: T22141

R.Kershaw, J.

submitted to the EMBL Data Library, September 1996

A.Reference number: Z19523

A.Accession: T22141

A.Status: preliminary; translated from GB/EMBL/DBJ

A.Molecule type: DNA

A.Residues: 1-466 <MIL>

A.Cross-References: EMBL:Z79755; PIDN:CAB02103.1; GSPDB:GN00019; CESP:F43G9.10

A.Experimental source: clone F43G9

C.Genetics:

A.Gene: CESP:F43G9.10

A.Map position: 1

A.Introns: 53/2; 316/3; 407/3

Query Match 10.8%; Score 141.5; DB 2; Length 466;

Best Local Similarity 25.5%; Pred. No. 0.27;

Matches 50; Conservative 42; Mismatches 69; Indels 35; Gaps 8;

```

QY 12 KRFVSEALDERRRKROEMKVRKPEDECEPEYDPSRYERLQOKRQOEYEE 71
Db 80 RRRRESDEEDRRRRRHHDYGRROVERPEVL-GKVEDESSSENEQSEDEDEKOEERRE 138
QY 72 OFKFNMY-----RGIDEFTNLFDEVSROOELIEKOR-----REELKELKRYRN 117
Db 139 RARMRLELHNNRREDEQ---DSAESDEDFFERRQMLDRATKREELK--REIK 193
QY 118 NKKVGIQENKKEVEK-----LTVKPIETKKNFSQAKLLAGAVKHSSEGSNV 167
Db 194 ELEEEDVEEESESESEDEDDDPVPLKPIETFRK--DRITLQAEKKEKE---I 248
QY 168 VKRIKPDPEPDKNQ 183
Db 249 LKRIEDEKRAEERRE 264

```

RESULT 15

T46481 hypothetical protein DKFp434A025.1 - human (fragment)

C.Species: Homo sapiens (man)

C.Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 04-Feb-2000

C.Accession: T46481

R.Duesterhoef, A.; Lauber, J.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.

submitted to the Protein Sequence Database, January 2000

A:Reference number: 223035
 A:Accession: T46481
 A>Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-1027 <AAA>
 A:Cross-references: EMBL:AL137755
 A:Experimental source: adult testis; clone DKFzp434A025
 C:Genetics:
 A:Note: DKFzp434A025.1

Query Match 10.88; Score 141; DB 2; Length 1027;
 Best Local Similarity 23.3%; Pred. No. 0.66;
 Matches 60; Conservative 36; Mismatches 79; Indels 82; Gaps 10;

```

QY 21 LDERRRRQOE---EMEVRKPEDEECPEEVDPRSL-----54
DB 45 IDRTKKKRGKDETEYSGSEEEVPEQEGEPSSIVNPGESTLRDPLRLQOEENKE 104
QY 55 -----YERLQEQKDRKQOEYEQF-----KFKNMVRGBLDEDETFLDEYSRQOE 98
DB 105 RSEALRQQLQEQQLRQOEYKRLAERQKRIEQCKEQRRRLQEQOR-EREARRQOE 163
QY 99 LIEKQRRREE--LKELEKYNLKKVGISQENKKEVEKKLV--KPIETKKEF-----147
DB 164 REQRRREOEKRRLEELERRRRREERREARRAEEKRRVEREQEYIRQLEEQRHLEVLQ 223
QY 148 ----SOAKLLAGAVKHKSSSESGNSVKRLKP-----PEPDDKNQEPSS 186
DB 224 QLOEQAMLL--ECRWREMEERHQAERLQRLQEQQAYLLSLQHDHRRPHPOHSQQPPPP 281
QY 187 CKSLGNTSLSGPSIICP 203
DB 282 ----QQERSKKPSFHAP 293
  
```

Search completed: February 16, 2001, 17:06:34
 Job time: 87 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2001, 17:05:07 ; Search time 10.27 Seconds
(Without alignments) 798.705 Million cell updates/sec

Title: US-09-602-597-2
Perfect score: 1306
Sequence: 1 MDGDDGNLIKKRFVSEAE.....NATKIVSSIFRNTFLPAP 254

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 88757 seqs, 32294092 residues

Total number of hits satisfying chosen parameters: 88757

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	158	12.1	1240	1 YN1_YEAST	P53935 saccharomyc
2	153	11.7	707	1 DREB_RAT	007266 rattus norv
3	151	11.6	649	1 DREB_HUMAN	016643 homo sapien
4	142.5	10.9	1178	1 MNNA_YEAST	P36044 saccharomyc
5	142	10.9	1407	1 TRHY_RABIT	P37709 oryctolagus
6	137	10.5	652	1 DREB_CHICK	P18302 gallus gall
7	136.5	10.5	678	1 GARP_PLAIF	P13816 plasmodium
8	132	10.1	405	1 TRT_CAEEL	027371 caenorhabd
9	131.5	10.1	1898	1 TRHY_HUMAN	007283 homo sapien
10	130.5	10.0	1403	1 YDEF3_SCHPO	010475 schizosacch
11	130	10.0	241	1 YOB8_CAEEL	P34623 caenorhabd
12	130	10.0	550	1 BLSA_HUMAN	002832 homo sapien
13	130	10.0	695	1 XE7_HUMAN	002040 homo sapien
14	128	9.8	805	1 IF2_AQUAE	067825 aquifex aeo
15	127.5	9.8	600	1 LAM2_CHICK	P14732 gallus gall
16	127.5	9.8	793	1 CALD_HUMAN	005682 homo sapien
17	126.5	9.7	407	1 M21_STRPY	P50468 streptococc
18	126.5	9.7	877	1 INCE_CHICK	P53352 gallus gall
19	126.5	9.7	2116	1 MYG2_DICDI	P08799 dictyostell
20	126	9.6	410	1 YG31_YEAST	P53866 saccharomyc
21	125.5	9.6	1959	1 MYSN_CHICK	P14105 gallus gall
22	125	9.6	897	1 EP15_MOUSE	P42567 mus musculu
23	125	9.6	1549	1 TRHY_SHEEP	P22793 ovis aries
24	124.5	9.5	657	1 KNOB_PLAIFN	P06719 plasmodium
25	123.5	9.5	816	1 HUNB_DROVI	P13361 drosophila
26	123	9.4	1234	1 PI33_MOUSE	P51432 mus musculu
27	122.5	9.4	771	1 CALD_CHICK	P12657 gallus gall
28	122.5	9.4	1085	1 YARA_SCHPO	009863 schizosacch
29	122.5	9.4	1220	1 IFP2_HUMAN	O60841 homo sapien
30	122.5	9.4	1391	1 MYG2_DROHY	O08696 drosophila
31	122	9.3	412	1 YNPI_CAEEL	P34554 caenorhabd
32	121.5	9.3	304	1 CEC1_CAEEL	P34518 caenorhabd
33	121.5	9.3	634	1 KNOB_PLAIFG	P09346 plasmodium

34	121.5	9.3	887	1 YLX8_CAEEL	P46504 caenorhabd
35	121.5	9.3	1023	1 GPT_DROME	P3438 drosophila
36	121.5	9.3	1118	1 UBP8_HUMAN	P40818 homo sapien
37	121.5	9.3	1940	1 MYSE_RAT	P12847 rattus norv
38	120.5	9.2	848	1 NFW_MOUSE	P08553 mus musculu
39	120.5	9.2	1940	1 MYSE_HUMAN	P11055 homo sapien
40	120	9.2	407	1 IE68_HSVSA	O01042 herpesvirus
41	119.5	9.2	396	1 TRT_DROME	P19351 drosophila
42	119.5	9.2	1790	1 USOL_YEAST	P25386 saccharomyc
43	119.5	9.2	1960	1 MTSN_HUMAN	P35379 homo sapien
44	119	9.1	285	1 TPM1_DROME	P06754 drosophila
45	118.5	9.1	542	1 TUL1_HUMAN	O00294 homo sapien

ALIGNMENTS

RESULT	1	STANDARD	PRT	1240 AA.
YN1_YEAST	YN1_YEAST			
AC	P53935			
DT	01-OCT-1996 (Rel. 34, Created)			
DT	01-OCT-1996 (Rel. 34, Last sequence update)			
DT	01-OCT-1996 (Rel. 34, Last annotation update)			
DE	HYPOTHETICAL 141.5 KDA PROTEIN IN YPT53-RHO2 INTERGENIC REGION.			
GN	YNL091W OR N2231.			
OS	Saccharomyces cerevisiae (Baker's yeast).			
OC	Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;			
OC	Saccharomycetaceae; Saccharomyces.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	SPRAIN-5288C / FY1679;			
RX	MEDLINE-96367601; PubMed-8771715;			
RA	Garcia-Cantalejo J.M., Boskovic J., Jimenez A.;			
RT	"Sequence analysis of a 14.2 kb fragment of Saccharomyces cerevisiae			
RT	chromosome XIV that includes the ypt53, tnnaleu and gsr m2 genes and			
RT	four new open reading frames."			
RL	Yeast 12:599-608(1996).			
CC	-1 SIMILARITY: TO S.POMBE SPAC29B6.10C.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL: X85811; CA59826.1; -			
DR	EMBL: Z71367; CA95967.1; -			
DR	HSSP: P03069; YISA.			
DR	SDD: S0005035; YNL091W.			
KW	Hypothetical protein.			
FT	DOMAIN 756..761			
SO	SEQUENCE 1240 AA; 141513 MW; 3FE9D265822D5778 CRC64: 4			
Query Match	12.1%; Score 158; DB 1; Length 1240;			
Best Local Similarity	23.8%; Pred. No. 0.051;			
Matches	54; Conservative 48; Mismatches 79; Indels 46; Gaps 8;			
QY	6 DGNLIKKRFVSEAELEDERKKRROEWEKYPKPPDEPECEVEYDPSLYERLOEQKDK 65			
DB	621 DNN---RLKLLDELE-EKKRRKREKKOKKREKKE-----KKRLQOLAAKEEK 669			
QY	66 QOEVEEEOFF-----KFKNNVRGDEDEETNPLDEVSQOELEIKOR-----104			
DB	670 REEKEKLELEERERKREARQKKYEAQKKKDEDERKKRLLEQQRREKQEKORQKE 729			
QY	105 -----REELEKEKERNRNKLVKGVISOENKKEVEKKLTVPRIETKKNFSQAQKLAVKH 159			
DB	730 ELKRRKEEKRIRE-QKRLQEKLOKEKEDEEERQRLIADALRKQKLNBEQTSANITLSA 788			

QY 160 KSSSEGSNVRKRLKDPEDPDKN-OEPSSC-----KSLGNTSLSGP 198
DB 789 KPTENGNGNPVSSSHPMNTNYOEDNCSINDEILKMNVSAAKSP 835

RESULT 2

REB_HUMAN STANDARD; PRT; 707 AA.
ID DREB_RAT
AC 007266;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE DREBRIN A.
GN DBNL.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-WISTAR: TISSUE-BRAIN, AND HIPPOCAMPUS;
RX MEDLINE=92305233; PubMed=1611026;
RA Shirao T., Obara N., Obara K.;
RT "Cloning of drebrin A and induction of neurite-like processes in
drebrin-transfected cells."
RL Neuroreport 3:109-112(1992).
CC -1- FUNCTION: DREBRINS MIGHT PLAY SOME ROLE IN CELL MIGRATION,
EXTENSION OF NEURONAL PROCESSES AND PLASTICITY OF DENDRITES,
RESPECTIVELY. BINDS ACTIN.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -1- ALTERNATIVE PRODUCTS: MULTIPLE FORMS OF DREBRINS RESULT FROM
ALTERNATIVE SPLICING OF THE SINGLE DREBRIN GENE DURING NEURAL
DEVELOPMENT.
CC -1- TISSUE SPECIFICITY: BRAIN NEURONS.
CC -1- MISCELLANEOUS: DREBRINS ARE CLASSIFIED INTO TWO FORMS OF THE
EMBRYONIC TYPE (E1 AND E2) AND ONE FORM OF THE ADULT TYPE (A). THE
TIME COURSE OF THEIR APPEARANCE ARE DIFFERENT FROM EACH OTHER.
CC THEIR STRUCTURES ARE CLOSELY RELATED.
CC CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X59267; CAA1957.1; -
DR INTERPRO: IPR002108; -
DR PFM: PF00241; Cofilin_ADF.1.
KW Actin-binding; Brain; Neurone; alternative splicing.
SQ SEQUENCE 707 AA; 77471 MW; B5279BFEB7B80AA CRC64;

Query Match 11.7%; Score 153; DB 1; Length 707;
Best Local Similarity 25.7%; Pred. No. 0.053;
Matches 61; Conservative 47; Mismatches 81; Indels 48; Gaps 10;

QY 17 SEAELEDRKRROEWEVVRKPEDEPCPEYVYPRSLYERLOQDKOQOYEEOQFK 76
DB 166 TDAVEEMKIRKREOFWEQAK---EEELRKEDEERKALDARLRFEOEMEDEROQ----- 218
QY 77 NMVGLDEDTNPLDEVSRQOELIEKORREBELKELKEYRNLLKVGISQENKKEVEK 136
DB 219 -----EERERRYRE--REQO-IEEHRKROQSLAEAEAKRRLKQOSIFGQDRDEE--- 265
QY 137 TVKPLETKNKSQAKLLAGAVKHKSSSEGSNVKRLKPD-PEPDDKNQPSCKSLGNTSL 195
DB 266 -----ESQMKKSESEV-----EEAAIIAQRPDNPREFFRQOQERVASASGSCDA 310
QY 196 SGPSTHCPASAIVCIGILPGLAY--SGSSDSSESSDSSE-----TINATGIYSSI 244
DB 311 PSPFNHRGPRYC-----PFIKASDSGSPSSSSSSSSSPPTPTTCHRTNLSL 363

RESULT 3

REB_HUMAN STANDARD; PRT; 649 AA.
ID DREB_HUMAN
AC Q16643;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE DREBRIN E.
GN DBNL.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-OSFROBLAST;
RA Fisher L.W., McBride O.W., Filipula D., Ibaraki K., Young M.F.;
RT "Human drebrin: cDNA sequence, mRNA tissue distribution and
chromosomal localization."
RL Neurosci. Res. Commun. 14:35-42(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE-FETAL BRAIN;
RX MEDLINE=94030036; PubMed=8216329;
RA Toda M., Shirao T., Minoshima S., Shimizu N., Toya S., Uyemura K.;
RT "Molecular cloning of cDNA encoding human drebrin E and chromosomal
mapping of its gene."
RL Biochem. Biophys. Res. Commun. 196:468-472(1993).
CC -1- FUNCTION: DREBRINS MIGHT PLAY SOME ROLE IN CELL MIGRATION,
EXTENSION OF NEURONAL PROCESSES AND PLASTICITY OF DENDRITES,
RESPECTIVELY. BINDS ACTIN.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -1- TISSUE SPECIFICITY: BRAIN NEURONS. ALSO FOUND IN THE HEART,
PLACENTA, SKELETAL MUSCLE, KIDNEY AND PANCREAS.
CC CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U00802; AAA16256.1; -
DR EMBL: D17530; BAA04480.1; -
DR MIM: 126660; -
DR INTERPRO: IPR002108; -
DR PFM: PF00241; Cofilin_ADF.1.
KW Actin-binding; Brain; Neurone.
SQ SEQUENCE 649 AA; 71425 MW; A1730E7C5FC32D88 CRC64;

Query Match 11.6%; Score 151; DB 1; Length 649;
Best Local Similarity 25.8%; Pred. No. 0.063;
Matches 49; Conservative 44; Mismatches 61; Indels 36; Gaps 7;

QY 17 SEAELEDRKRROEWEVVRKPEDEPCPEYVYPRSLYERLOQDKOQOYEEOQFK 76
DB 166 TDAVEEMKIRKREOFWEQAK---EEELRKEDEERKALDARLRFEOEMEDEROQ----- 218
QY 77 NMVGLDEDTNPLDEVSRQOELIEKORREBELKELKEYRNLLKVGISQENKKEVEK 136
DB 219 -----EERERRYRE--REQO-IEEHRKROQSLAEAEAKRRLKQOSIFGQDRDEE--- 265
QY 137 TVKPLETKNKSQAKLLAGAVKHKSSSEGSNVKRLKPD-PEPDDKNQPSCKSLGNTSL 195
DB 266 -----ETHMKKSESEV-----EEAAIIAQRPDNPREFFRQOQERVASASGSCDV 310
QY 196 SGPSTHCPSA 205
DB 311 PSPFNHRGCS 320

```

RESULT 4
MNN4_YEAST STANDARD: PRT; 1178 AA.
ID MNN4_YEAST
AC P36044; P36043; P89095;
DT 01-JUN-1994 (Rel. 29, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE MNN4 PROTEIN.
GN MNN4 OR YKL200C/YKL201C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
NC Saccharomycetaceae; Saccharomyces.
RC [1]
RP SEQUENCE FROM N.A.
RC STRAIN-S288C.
RX MEDLINE-97175967; PubMed-9023541;
RA Odani T., Shima Y.-I., Tanaka A., Jigami Y.;
RT "Cloning and analysis of the MNN4 gene required for phosphorylation
RT of N-linked oligosaccharides in Saccharomyces cerevisiae.";
RL Glycobiology 6:805-810(1996).
RN [2]
RP SEQUENCE FROM N.A.
RA Maia E Silva A., Bossier P., Vilela C., Fernandes L., Soares H.,
RA Guerreiro P., Rodrigues-Pousada C.;
RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: MAY FUNCTION AS A POSITIVE REGULATOR FOR
CC MANNOSYLPHOSPHATE TRANSFERASE. IS REQUIRED TO MEDIATE
CC MANNOSYLPHOSPHATE TRANSFER IN BOTH THE CORE AND OUTER CHAIN
CC PORTIONS OF N-LINKED OLIGOSACCHARIDES.
CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN (PROBABLE).
CC -1- SIMILARITY: TO YEAST YJR061W.
CC -1- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO
CC FRAMESHIFTS, ONE OF WHICH PRODUCES TWO SEPARATE ORFS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: D83006; BAAL1676.1; -
DR EMBL: Z28201; CAAB2046.1; -
DR EMBL: Z28200; CAAB2044.1; -
DR PIR: S38037; S38037.
DR PIR: S38038; S38038.
DR YEPD: 6530; -
DR SGD: S0001683; MNN4.
KW Transmembrane; Signal-anchor; Repeat.
FT DOMAIN 1 27 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 28 48 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT FT (POTENTIAL).
FT DOMAIN 49 1178 LUMENAL (POTENTIAL).
FT DOMAIN 1032 1174 ARG/GLU/LYS-RICH (HIGHLY CHARGED).
FT DOMAIN 1042 1174 17 x 8 AA TANDEM REPEATS OF K-K-K-E-E-
FT FT E-E.
FT REPEAT 1042 1049 1.
FT REPEAT 1050 1057 2.
FT REPEAT 1058 1065 3.
FT REPEAT 1066 1073 4.
FT REPEAT 1074 1081 5.
FT REPEAT 1082 1089 6.
FT REPEAT 1090 1097 7 (APPROXIMATE).
FT REPEAT 1098 1105 8.
FT REPEAT 1106 1113 9 (APPROXIMATE).
FT REPEAT 1114 1121 10 (APPROXIMATE).
FT REPEAT 1122 1129 11 (APPROXIMATE).
FT REPEAT 1130 1137 12.
FT REPEAT 1138 1144 13 (APPROXIMATE).
FT REPEAT 1145 1152 14 (APPROXIMATE).
FT REPEAT 1153 1160 15 (APPROXIMATE).
FT REPEAT 1161 1168 16 (APPROXIMATE).

```

```

FT REPEAT 1169 1174 17 (INCOMPLETE).
FT DOMAIN 37 40 POLY-TLE.
FT DOMAIN 1042 1045 POLY-LYS.
FT DOMAIN 1046 1049 POLY-GLU.
FT DOMAIN 1050 1053 POLY-LYS.
FT DOMAIN 1054 1057 POLY-LYS.
FT DOMAIN 1058 1061 POLY-LYS.
FT DOMAIN 1062 1065 POLY-LYS.
FT DOMAIN 1066 1069 POLY-LYS.
FT DOMAIN 1070 1073 POLY-GLU.
FT DOMAIN 1074 1077 POLY-LYS.
FT DOMAIN 1078 1081 POLY-GLU.
FT DOMAIN 1082 1085 POLY-LYS.
FT DOMAIN 1086 1089 POLY-GLU.
FT DOMAIN 1094 1097 POLY-GLU.
FT DOMAIN 1098 1101 POLY-LYS.
FT DOMAIN 1098 1101 POLY-LYS.
FT DOMAIN 1102 1105 POLY-GLU.
FT DOMAIN 1134 1137 POLY-GLU.
FT DOMAIN 1157 1160 POLY-GLU.
FT DOMAIN 1165 1168 POLY-GLU.
SQ SEQUENCE 1178 AA; 139380 MW; BC05DAE0AEFCB282 CRC64;

Query Match 10.9%; Score 142.5; DB 1; Length 1178;
Best Local Similarity 27.3%; Pred. No. 0.35; Mismatches 61; Indels 37; Gaps 6;
Matches 51; Conservative 38;

QY 6 DGNLIIRKRVSEALDERRRKQ---EEWKVRKPEDPECEPEVYDPSIYERLQEQ 61
DB 1018 DKPIIIVYEDYAAKLLLEKRRRRKKEEKKKEEKKKEE-----EKKKKE 1070
QY 62 KDRKQOEYEROFKKNVNRGLDETFVLEVSNOQLIKORREELKEKERNMLK 121
DB 1071 EEEKKKEEKKKEE-----EEKKKOEKKKEEKKKEEKKKEGEGEKMN 1117
QY 122 VGISOENKKVEKKLVKPIETKNNKFSQAKILGAVYKKSSEGSNSYKRLKPPDPDKN 181
DB 1118 E--DEKKKNEDEKKKNEEKKKQDE-----KNKNE--DEKKKOEKKKKN 1164
QY 182 QEPSSCK 188
DB 1165 EEEKKK 1171

RESULT 5
TRHY_RABIT STANDARD: PRT; 1407 AA.
AC P37709;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE TRICHOHYALIN.
GN THH.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
RN [1]
RP SEQUENCE FROM N.A.
RA Fietz M.J., Rogers G.E.;
RL Submitted (DEC-1992) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: INTERMEDIATE FILAMENT-ASSOCIATED PROTEIN THAT ASSOCIATES
CC IN REGULAR ARRAYS WITH KERATIN INTERMEDIATE FILAMENTS (KIF) OF THE
CC INNER ROOT SHEATH CELLS OF THE HAIR FOLLICLE AND THE GRANULAR
CC LAYER OF THE EPIDERMIS. IT LATER BECOMES CROSS-LINKED TO KIF BY
CC ISODIPEPTIDE BONDS. IT MAY SERVE AS SCAFFOLD PROTEIN, TOGETHER
CC WITH INVOLUCRIN, IN THE ORGANIZATION OF THE CELL ENVELOPE OR EVEN
CC ITS OWN CELL ENVELOPE TO THE KIE NETWORK. IT MAY BE INVOLVED IN
CC ITS OWN CALCIUM-DEPENDENT POSTSYNTHETIC PROCESSING DURING TERMINAL
CC DIFFERENTIATION.
CC -1- SUBUNIT: HOMODIMER (PROBABLE).
CC -1- TISSUE SPECIFICITY: FOUND IN THE HARD KERATINIZING TISSUES SUCH AS
CC THE INNER ROOT SHEAT (IRS) OF HAIR FOLLICLES AND MEDULLA, AND IN
CC THE FILIFORM PAPILLAE OF DORSAL TONGUE EPITHELIUM (PROBABLE).

```

```
CC -1- DEVELOPMENTAL STAGE: EXPRESSED DURING LATE DIFFERENTIATION OF THE EPIDERMIS.
CC CC
CC -1- DOMAIN: CONSISTS OF NINE DOMAINS. DOMAIN 1 CONTAINS TWO EF-HAND CALCIUM-BINDING DOMAINS. DOMAINS 2-4, 6, AND 8 ARE ALMOST ENTIRELY ALPHA-HELICAL, CONFIGURED AS A SERIES OF PEPTIDE REPEATS OF VARYING REGULARITY, AND ARE THOUGHT TO FORM A SINGLE-STANDED ALPHA-HELICAL ROD STABILIZED BY IONIC INTERACTIONS. DOMAIN 6 IS THE MOST REGULAR AND MAY BIND KIP DIRECTLY BY IONIC INTERACTIONS. DOMAINS 5 AND 7 ARE LESS WELL ORGANIZED AND MAY INDUCE FOLDS IN THE MOLECULE. DOMAIN 9 CONTAINS THE C-TERMINUS, CONSERVED AMONG DIFFERENT SPECIES.
CC CC
CC -1- PTM: KNOWN SUBSTRATE OF TRANSGLUTAMINASE. SOME 200 ARGININES ARE PROBABLY CONVERTED TO C-TERMINAL SERINES BY PEPTIDYLARGININE DEIMINASE.
CC CC
CC -1- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE S-100 FAMILY.
CC CC
CC -1- SIMILARITY: CONTAINS 2 EF-HAND CALCIUM-BINDING DOMAINS.
-----
CC CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce/ or send an email to license@sib-sib.ch).
CC CC
CC -----
CC DR EMBL; Z19092; CAA79519.1; -.
CC DR PIR; S28589; S28589.
CC DR HSSP; P02633; 1BQC.
CC DR INTERPRO; IPRO001751; -.
CC DR INTERPRO; IPR002048; -.
CC DR PFAM; PF01023; S_100; 1.
CC DR PFAM; PF00036; efhand; 1.
CC DR PROSITE; PS00018; EF_HAND; 1.
CC DR PROSITE; PS00303; S100_CABP; 1.
CC KW Repeat; Calcium-binding.
CC FT DOMAIN 1 91 S-100 LIKE.
CC FT CA_BIND 22 33 SITE I (LOW AFFINITY) (POTENTIAL).
CC FT CA_BIND 62 73 SITE II (HIGH AFFINITY) (POTENTIAL).
CC SQ SEQUENCE 1407 AA; 183781 MW; AE17D2A159F12B7F CRC64;

Query Match 10.9%; Score 142; DB 1; Length 1407;
Best Local Similarity 26.9%; Pred. No. 0.46; Mismatches 48; Indels 46; Gaps 8;
Matches 49; Conservative 39;

QY 13 KRYSSEALDERKKR-----QEWEKV-----RKPEDPECPREYYDPRLSYRL 58
   |||::||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 215 KRELKEEE--QORRREREQHERALQEEBOQLLRQRWRREPREEQQOLRRLEIEIREREO 273
   :::::|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY 59 QEQKRRKQ-----EVEE-QEFKMMVRGLDEDFNLFDEVSRQOEELIKQRREEL 109
   :::::|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 274 EQEERREOQLRREOREQERREOQLRRELEIEIRERERORLEQERREORLEQERREOQL 333
   :::::|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY 110 R-ELKEYNNLKLVISOENKKEVKKLTVPRIETFKNKFSQAKLACGAVKRKSSSGNSV 168
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 334 KRELKEIEIRERGR--LQEEERRE-----QLLAEEVREDARRARGESTL 372
   :::::|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY 169 KR 170
   |
Db 373 TR 374

RESULT 6
DREB_CHICK STANDARD; PRT; 652 AA.
AC PI8302; Q91358; Q91359;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE DREBRINS A, EI AND E2.
DB ENI.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```

OC Archaeosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
CC Gallus.
RN [1]
RP SEQUENCE FROM N.A. (FORMS E1 AND E2).
RC TISSUE-BRAIN:
RX MEDLINE=89089137; PubMed=3208110;
RA Kojima N., Kato Y., Shirao T., Ohta K.;
RT "Molecular cloning of a developmentally regulated brain protein,
RT chicken drebrin A and its expression by alternative splicing of the
RT drebrin gene.";
RL Brain Res. 101. Brain Res. 19:101-114(1993).
RN [2]
RP SEQUENCE FROM N.A. (FORM A).
RX MEDLINE=93368392; PubMed=8361332;
RA Kojima N., Shirao T., Ohta K.;
RT "Alternative splicing of a developmentally regulated brain protein,
RT chicken drebrin A and its expression by alternative splicing of the
RT drebrin gene.";
RL Brain Res. 101. Brain Res. 19:101-114(1993).
CC -1- FUNCTION: DREBRINS MIGHT PLAY SOME ROLE IN CELL MIGRATION,
CC EXTENSION OF NEURONAL PROCESSES AND PLASTICITY OF DENDRITES,
CC RESPECTIVELY.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -1- ALTERNATIVE PRODUCTS: MULTIPLE FORMS OF DREBRINS RESULT FROM
CC ALTERNATIVE SPLICING OF THE SINGLE DREBRIN GENE DURING NEURAL
CC DEVELOPMENT.
CC -1- TISSUE SPECIFICITY: BRAIN NEURONS.
CC -1- MISCELLANEOUS: DREBRINS ARE CLASSIFIED INTO TWO FORMS OF THE
CC EMBRYONIC TYPE (E1 AND E2) AND ONE FORM OF THE ADULT TYPE (A). THE
CC TIME COURSE OF THEIR APPEARANCE ARE DIFFERENT FROM EACH OTHER.
CC THEIR STRUCTURES ARE CLOSELY RELATED.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/).
CC or send an email to license@isb-sib.ch.
CC -----
DR EMBL; M36961; AAA48750.1; -.
DR EMBL; S65296; AAB28012.1; JOINED.
DR EMBL; S65279; AAB28012.1; JOINED.
DR EMBL; S65280; AAB28012.1; JOINED.
DR EMBL; S65281; AAB28012.1; JOINED.
DR EMBL; S65288; AAB28012.1; JOINED.
DR EMBL; S65289; AAB28012.1; JOINED.
DR EMBL; S65290; AAB28012.1; JOINED.
DR EMBL; S65291; AAB28012.1; JOINED.
DR EMBL; S65292; AAB28012.1; JOINED.
DR EMBL; S65294; AAB28012.1; JOINED.
DR EMBL; S65297; CAB32513.1; -.
DR EMBL; S65230; AAB28010.1; -.
DR PIR; A43776; A43776.
DR INTERPRO; IPR002108; -.
DR PFAM; PF00241; coffin_ADF.1.
KW Actin-binding; Brain; Neurone; Alternative splicing.
FT DOMAIN 47 50 POLY-GLY.
FT DOMAIN 332 339 POLY-SER.
FT DOMAIN 542 552 POLY-GLU.
FT DOMAIN 643 646 POLY-GLU.
FT VASPLC 317 362 MISSING (IN ISOFORM E1).
FT VASPLC 307 405 MISSING (IN ISOFORM E2).
FT VASPLC 552 552 E -> EE (IN REF. 1).
FT CONFLICT 552 552
SQ SEQUENCE 652 AA; 71535 MW; 8BF72CA549833EB5 CRC64;

Query Match 10.5%; Score 137; DB 1; Length 652;
Best Local Similarity 24.8%; Pred. No. 0.38;
Matches 54; Conservative 43; Mismatches 73; Indels 48; Gaps 8;

17 SEAEIDERRKRQOEWEKVRPEDECEPEEYVDRSLYERLQEOCKDRKQOEYEOFKRK 76
::: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 166 TDATVEMKRLNDEQWEOAK---EEELRKKEERKALDARLFEQWREDEORLQO--- 218

```

```

OY 77 NWVRGLDEDETNFLDEVSROOELIEKORREBELKEKEYNNLKVYISOENKEVEKLL 136
DB 219 -----EERRRRRE-----REBOIEHRRKQSMFAEAFORLEOSTIFEQOEDDDKQ 268
OY 137 TVAPLETNKNFSQAKLAGAVKHRSSESGNSVRLKPD-PEPDDKNOE---PSSCKSLGN 192
DB 269 LRR-----SESEV-----EEAATIAQRPNDRREFEKOQERVAASSGDAISP 310
OY 193 TSLSGPSIHCPSAVAVCIGILPGICANSGSSDSSSDS 230
DB 311 GSHRTGRLHCPFIKT-----ADSGPSSSSSSSS 339

RESULT 7
GARP_PLAUF STANDARD; PRT; 678 AA.
ID GARP_PLAUF STANDARD; PRT; 678 AA.
AC P13816;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, last sequence update)
DT 15-JUL-1999 (Rel. 38, last annotation update)
DE GLUTAMIC ACID-RICH PROTEIN PRECURSOR.
GN GARP.
OS Plasmodium falciparum (isolate FC27 / Papua New Guinea).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=69040048; PubMed=2903445;
RA Triglia T., Stahl H.-D., Crewther P.E., Silva A., Anders R.F.,
RT Kemp D.J.;
RT "Structure of a Plasmodium falciparum gene that encodes a glutamic
RL Mol. Biochem. Parasitol. 31:199-202(1988).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: J03998; AAA29605.1;
CC PIR: A54514; A54514.
DR Repeat; Malaria; Antigen; Signal.
KM Repeat; Malaria; Antigen; Signal.
FT CHAIN 1 25
FT DOMAIN 26 678 GLUTAMIC ACID-RICH PROTEIN.
FT DOMAIN 120 164 15 X 3 AA TANDEM REPEATS OF K-K-X.
FT DOMAIN 372 416 9 X APPROXIMATE TANDEM REPEATS.
FT DOMAIN 417 441 5 X APPROXIMATE TANDEM REPEATS.
FT DOMAIN 576 604 POLY-GLU.
FT DOMAIN 605 653 7 X APPROXIMATE TANDEM REPEATS.
FT DOMAIN 654 663 POLY-GLU.
SQ SEQUENCE 678 AA; 80551 MW; 2A8F85606496EA9E CRC64;

Query Match 10.5%; Score 136.5; DB 1; Length 678;
Best Local Similarity 26.5%; Pred. No. 0.43; Mismatches 33; Gaps 8;
Matches 57; Conservative 45; Indels 33; Gaps 8;
OY 12 KRFVSEAEIDERRRRROEWEKVR--KPEDPECEPEEYVDPISLYERLOEKDRKOQXY 69
DB 268 KEKEMKEOEIEKKKKKKOEKKKKOEKKKKOEKKKKOEKKKKOEKKKKOEKKKKOE 327
OY 70 EEOFKFRNMVRGLD-----EDETNF-----LDEVSRQOELIEKORREBELKEL 112
DB 328 KEKKKKKKHDEKNETMQOPDTSEETNNEIMVPLPFDVTPPEHKEGEHEEHEK- 386
OY 113 KEVRNNLKVYISOENKEVEKLLTVKPIETKKNFSQAKLAGAVKHRSSESGNSVRLK 170
DB 387 GEHKEGEHK--EEHKEEHEKKEHKEHKGKKGKDKGKDKGKDKGKDKGKDKGKDK 442
OY 171 LKPDPEDDKN-----QEPSSCKSLGNTSLSGP 198

```

```

DB 443 YKNVIEDEEDKDGVEIINLEDKKEACEBQHTIVESRP 477
RESULT 8
TRT CAEEL
ID TRT CAEEL STANDARD; PRT; 405 AA.
AC 027371;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, last sequence update)
DT 15-JUL-1998 (Rel. 36, last annotation update)
DE TROPONIN T.
GN MUP-2 OR T225.5.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2; TISSUE=MUSCLE;
RX MEDLINE=96176302; PubMed=8601585;
RA Myers C.D., Gon P.-Y., Allen T.S., Bucher E.A., Bogaert T.;
RT "Developmental genetic analysis of troponin T mutations in striated
RT and nonstriated muscle cells of Caenorhabditis elegans.";
RL J. Cell Biol. 132:1061-1077(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Minx P.;
RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: TROPONIN T IS THE TROPOMYOSIN-BINDING SUBUNIT OF
CC TROPONIN, THE THIN FILAMENT REGULATORY COMPLEX WHICH CONFERS
CC CALCIUM-SENSITIVITY TO STRIATED MUSCLE ACTOMYOSIN ATPASE ACTIVITY
CC (BY SIMILARITY).
CC
CC -1- SIMILARITY: TO VERTEBRATE SKELETAL AND CARDIAC TNT.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: U44759; AA86906.1;
CC DR EMBL: U43282; AA83615.1;
CC DR WORMPEP: T22E5.5; CE04994.
CC DR INTERPRO: IPR001978;
CC PFAM: PF00992; Troponin; 1.
FT DOMAIN 11 23 POLY-GLU.
FT DOMAIN 75 78 POLY-GLU.
SQ SEQUENCE 405 AA; 47041 MW; 1D3AC9EBB641C65B CRC64;

Query Match 10.1%; Score 132; DB 1; Length 405;
Best Local Similarity 25.9%; Pred. No. 0.44; Mismatches 65; Indels 26; Gaps 6;
Matches 44; Conservative 33; Mismatches 65; Indels 26; Gaps 6;
OY 17 SEAEIDERRRRROEWEKVRKPEDPECEPEEYVDPISLYER--LOEKDRKOQXYEOP 73
DB 9 SDEVEEEVEEETE--EAPAEAEPEPETTEEVAVAPPEVKRRAPVOEKKPAPMTAEITA 66
OY 74 KRFVSEAEIDEDTNFLDEVSROOELIEKORREBELKEKEYNNLKVYISOENKEVE 133
DB 67 MLAARRHHEEEFAKLID--YEQRVRLVKQIDELREKE-----KOEKRRAE 113
OY 134 KRLTVPIETNKNFSQAKLAGAVKHRSSESGNSVRLKPDPEDDKNOE 183
DB 114 RE-----EDERQFAERRRQDDERRRKEDE-----KKAADAKIKIKNEE 153

RESULT 9
TRHY_HUMAN STANDARD; PRT; 1898 AA.
ID TRHY_HUMAN

```

AC 007283;
 DT 01-OCT-1994 (Rel. 30, Created)
 DT 01-OCT-1994 (Rel. 30, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE TRICHOHYALIN.
 GN THH OR TRHY OR THL.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 NC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-93280194; PubMed-7685034;
 RA Lee S.-C., Kim I.-G., Marekov L.N., O'Keefe E.J., Parry D.A.D.,
 RA Steinert P.M.;
 RT "The structure of human trichohyalin. Potential multiple roles as a
 functional EF-hand-like calcium-binding protein, a cornified cell
 envelope precursor, and an intermediate filament-associated (cross-
 linking) protein.";
 RT J. Biol. Chem. 268:12164-12176(1993).
 RL (2)
 RN SEQUENCE OF 1731-1898 FROM N.A., AND CHARACTERIZATION.
 RP MEDLINE-93315897; PubMed-7686953;
 RX O'Keefe E.J., Hamilton E.H., Lee S.-C., Steinert P.M.;
 RA "Trichohyalin: a structural protein of hair, tongue, nail, and
 epidermis.";
 RT J. Invest. Dermatol. 101:655-715(1993).
 RL J. Invest. Dermatol. 101:655-715(1993).
 CC -1- FUNCTION: INTERMEDIATE FILAMENT-ASSOCIATED PROTEIN THAT ASSOCIATES
 IN REGULAR ARRAYS WITH KERATIN INTERMEDIATE FILAMENTS (KIF) OF THE
 INNER ROOT SHEATH CELLS OF THE HAIR FOLLICLE AND THE GRANULAR
 LAYER OF THE EPIDERMIS. IT LATER BECOMES CROSS-LINKED TO KIF BY
 13ODIPEPTIDE BONDS. IT MAY SERVE AS SCAFFOLD PROTEIN, TOGETHER
 WITH INVOLUCRIN, IN THE ORGANIZATION OF THE CELL ENVELOPE OR EVEN
 ANCHOR THE CELL ENVELOPE TO THE KIF NETWORK. IT MAY BE INVOLVED IN
 ITS OWN CALCIUM-DEPENDENT POSTSYNTHETIC PROCESSING DURING TERMINAL
 DIFFERENTIATION.
 CC -1- SUBUNIT: MONOMER (PROBABLE).
 CC -1- TISSUE SPECIFICITY: FOUND IN THE HARD KERATINIZING TISSUES SUCH AS
 THE INNER ROOT SHEAT (IRS) OF HAIR FOLLICLES AND MEDULLA, AND IN
 THE FLITIFORM PAPILLAE OF DORSAL TONGUE EPITHELIUM (PROBABLE).
 CC -1- DEVELOPMENTAL STAGE: EXPRESSED DURING LATE DIFFERENTIATION OF
 THE EPIDERMIS.
 CC -1- DOMAIN: CONSISTS OF NINE DOMAINS. DOMAIN 1 CONTAINS TWO EF-HAND
 CALCIUM-BINDING DOMAINS. DOMAINS 2-4, 6, AND 8 ARE ALMOST
 ENTIRELY ALPHA-HELICAL, CONFIGURED AS A SERIES OF PEPTIDE REPEATS
 OF VARYING REGULARITY, AND ARE THOUGHT TO FORM A SINGLE-STRAINED
 ALPHA-HELICAL ROD STABILIZED BY IONIC INTERACTIONS. DOMAIN 6 IS
 THE MOST REGULAR AND MAY BIND KIF DIRECTLY BY IONIC INTERACTIONS.
 CC DOMAINS 5 AND 7 ARE LESS WELL ORGANIZED AND MAY INDUCE FOLDS IN
 THE MOLECULE. DOMAIN 9 CONTAINS THE C-TERMINUS, CONSERVED AMONG
 DIFFERENT SPECIES.
 CC -1- PTM: KNOWN SUBSTRATE OF TRANSGLUTAMINASE. SOME 200 ARGININES ARE
 PROBABLY CONVERTED TO CITRULLINES BY PEPTIDYLARGININE DEIMIDASE.
 CC -1- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE S-100
 FAMILY.
 CC -1- SIMILARITY: CONTAINS 2 EF-HAND CALCIUM-BINDING DOMAINS.
 CC CC
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: L09190; AAA65582.1; -
 DR PIR: A45973; A45973.
 DR HSSP: P02633; IB0C.
 DR MIM: 190370; -
 DR INTERPRO: IPR001751; -
 DR INTERPRO: IPR002048; -
 DR PFAM: PF01023; S_100; 1.
 DR PFAM: PF00036; efhand; 1.
 DR PROSITE: PS00018; EF_HAND; 1.

DR PROSITE; PS00303; S100_CABP; 1.
 KW Repeat; Calcium-binding.
 FT DOMAIN 1 91
 FT CA_BIND 22 33
 FT CA_BIND 62 73
 FT DOMAIN 314 390
 FT REPEAT 314 326
 FT REPEAT 327 339
 FT REPEAT 340 351
 FT REPEAT 352 364
 FT REPEAT 365 377
 FT REPEAT 378 390
 FT DOMAIN 391 444
 FT REPEAT 391 396
 FT REPEAT 397 402
 FT REPEAT 403 408
 FT REPEAT 409 414
 FT REPEAT 415 420
 FT REPEAT 421 426
 FT REPEAT 427 432
 FT REPEAT 433 438
 FT REPEAT 439 444
 FT DOMAIN 444 702
 FT DOMAIN 923 1162
 FT REPEAT 923 952
 FT REPEAT 953 982
 FT REPEAT 983 1012
 FT REPEAT 1013 1042
 FT REPEAT 1043 1072
 FT REPEAT 1073 1102
 FT REPEAT 1103 1132
 FT REPEAT 1133 1162
 FT DOMAIN 1250 1849
 FT CONFLICT 1752 1752
 FT CONFLICT 1794 1801
 FT CONFLICT 1857 1857
 FT CONFLICT 1880 1880
 SQ SEQUENCE 1898 AA; 247219 MW; A7AB5947FB62E31D CRC64;
 Query Match 10.18; Score 131.5; DB 1; Length 1898;
 Best Local Similarity 25.88; Pred. No. 2.4;
 Matches 41; Conservative 35; Mismatches 46; Indels 37; Gaps 5;
 QY 12 KKRFSSEAL-----DERRKRQEWKVRKPED-----PECEEYDPSLYER 57
 DB 1004 EKYREBELQOEEOQLREERERKRROEWERYKKDELQOEEOQLREERERKRLOER 1063
 QY 58 -----LQEKD-----RKQOEYEOEFKMMVRLGDEDTNFIDEVSRQOE 99
 DB 1064 ERYREBELQOEEOQLGEBRETRRQELRYRKKEELOOEEOQLREPERKRROER 1123
 QY 100 TEKORREELKE-----LKEYRNLIKVGISOENKKEVE 133
 DB 1124 EROCREEELOOEEOQLREERERKRROELEROYREEE 1162
 RESULT 10
 YDF3_SCHPO STANDARD; PRT: 1403 AA.
 AC 010475;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE PROBABLE EUKARYOTIC INITIATION FACTOR C17C9.03.
 GN SPAC17C9.03.
 OS Schizosaccharomyces pombe (fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetales;
 CC Schizosaccharomycetaceae; Schizosaccharomycetes.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-972;

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2001, 17:05:07 ; Search time 13.72 Seconds
(without alignments)
332.441 Million cell updates/sec

Title: US-09-602-597-2
Perfect score: 1306
Sequence: 1 MDGGDGNLIKRRFVSEAE.....NATGKIVSIPRTWFLFAP 254

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 174772 seqs, 17957048 residues

Total number of hits satisfying chosen parameters: 174772

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
1: /cgn2_6/prodata/2/1aa/5A.COMB.pep:*
2: /cgn2_6/prodata/2/1aa/5B.COMB.pep:*
3: /cgn2_6/prodata/2/1aa/6.COMB.pep:*
4: /cgn2_6/prodata/2/1aa/PCITUS.COMB.pep:*
5: /cgn2_6/prodata/2/1aa/Backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	133	10.2	1162	2	US-08-728-323A-2
2	131.5	10.1	1898	1	US-08-056-200-94
3	131.5	10.1	1898	2	US-08-800-644-94
4	125	9.6	897	1	US-08-095-737-4
5	125	9.6	897	1	US-08-480-145-4
6	125	9.6	897	2	US-08-477-388-4
7	123.5	9.5	1507	3	US-08-929-328-5
8	120.5	9.2	542	1	US-08-701-380-2
9	120.5	9.2	542	3	US-09-032-365A-13
10	120	9.2	432	3	US-08-933-750C-47
11	120	9.2	432	3	US-09-234-613-47
12	119	9.1	285	1	US-08-149-809-24
13	118	9.0	1588	4	PCR-US93-07261-11
14	118	9.0	1663	4	PCR-US93-07261-16
15	116	8.9	411	2	US-08-741-134-6
16	115	8.8	532	1	US-08-285-440-5
17	115	8.8	532	1	US-08-630-349-5
18	114.5	8.8	905	2	US-08-574-959A-9
19	114.5	8.8	1135	2	US-08-574-959A-7
20	113.5	8.7	717	3	US-08-910-925-1
21	111.5	8.5	700	1	US-07-720-589-2
22	111.5	8.5	700	2	US-08-785-190-2
23	111.5	8.5	700	4	PCR-US92-05539-2
24	110	8.4	290	2	US-08-903-801-1
25	109.5	8.4	703	3	US-08-910-925-4
26	109.5	8.4	2182	2	US-08-487-826B-16
27	109.5	8.4	3135	1	US-08-323-170B-2
28	108	8.3	424	2	US-08-951-148-9

29	108	8.3	424	2	US-09-165-234-9	Sequence 9, Appli
30	108	8.3	424	3	US-09-274-570-9	Sequence 9, Appli
31	107.5	8.2	558	1	US-08-285-440-6	Sequence 6, Appli
32	107.5	8.2	558	1	US-08-630-349-6	Sequence 6, Appli
33	106.5	8.2	224	2	US-08-272-255-16	Sequence 16, Appli
34	106.5	8.2	224	4	PCR-US95-08565-16	Sequence 6, Appli
35	106	8.1	743	2	US-08-795-475-6	Sequence 6, Appli
36	106	8.1	743	3	US-08-990-114-3	Sequence 3, Appli
37	105.5	8.1	714	2	US-09-149-934-4	Sequence 3, Appli
38	105.5	8.1	746	3	US-08-242-932-2	Sequence 2, Appli
39	105.5	8.1	984	1	US-08-714-481-2	Sequence 2, Appli
40	105.5	8.1	984	4	PCR-US95-06111-2	Sequence 2, Appli
41	105.5	8.1	984	4	US-08-312-949-4	Sequence 4, Appli
42	105	8.0	288	3	US-08-446-201-4	Sequence 4, Appli
43	105	8.0	288	3	US-08-465-746-2	Sequence 4, Appli
44	105	8.0	619	1	US-08-214-164-2	Sequence 2, Appli
45	105	8.0	619	1	US-08-214-164-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
; Sequence 2, Application US/08728323A
; Patent No. 5948676
; GENERAL INFORMATION:
; APPLICANT: Chang, Yuan
; APPLICANT: Bohenzky, Roy A.
; APPLICANT: Russo, James J.
; APPLICANT: Egelman, Isidore S.
; APPLICANT: Moore, Patrick S.
; TITLE OF INVENTION: Immediate Early Protein From Kaposi's
; TITLE OF INVENTION: Sarcoma-Associated Herpesvirus, DNA
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/728,323A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-278-0400
; TELEFAX: 212-391-0525
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1162 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-728-323A-2

Query Match 10.2%; Score 133; DB 2; Length 1162;
Best Local Similarity 25.5%; Pred. No. 0.0014;
Matches 41; Conservative 41; Mismatches 61; Indels 18; Gaps 4;
QY 18 EAELDERRRKRRQENWEKVKRPEDPECEPEYVDPSLVERLQEQDKRQOEYEDQFKFN 77

Db 836 EQLEEEVEVEBOEVEBOEBOEBOELEEV-----BOEBOEBOEBOELE----- 883
QY 78 MVRGLDEETNFDLVSQOELIKORREBELKEKYRNNLKKYISQENKKEVEKLT 137
Db 884 -----VEEBOEBOELEVEVEBOEBOELEVEBO-----EQGVBQOEOETVEEPTI 938
QY 138 VKPIETKNKFS-QAKLAGAVKHKSSGNSVKRLKPPPEP 177
Db 939 LHSSSEDEMEVDYVSTHQIASSPGDMPDDDPQGP 979

RESULT 2
US-08-056-200-94
; Sequence 94, Application US/08056200
; Patent No. 5616500
; GENERAL INFORMATION:
; APPLICANT: Steinert, Peter M.
; APPLICANT: Lee, Seung-Chul
; APPLICANT: Kim, In-Gyu
; APPLICANT: Chung, Soo-Il
; APPLICANT: Park, Sang-Chul
; TITLE OF INVENTION: Trichohyalin and Transglutaminase-3 and
; NUMBER OF SEQUENCES: 117
; TITLE OF INVENTION: Methods of Using Same
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe, Martens, Olson & Bear
; STREET: 620 Newport Center Drive, Sixteenth Floor
; CITY: Newport Beach
; STATE: CA
; COUNTRY: U.S.A.
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/056, 200
; FILING DATE: 30-APR-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Fedrick, Michael F.
; REGISTRATION NUMBER: 36,799
; REFERENCE/DOCKET NUMBER: NIH054.001A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (714) 760-0404
; TELEFAX: (714) 760-9502
; INFORMATION FOR SEQ ID NO: 94:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1898 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-056-200-94

Query Match 10.1%; Score 131.5; DB 1; Length 1898;
Best Local Similarity 25.8%; Pred. No. 0.0039;
Matches 41; Conservative 35; Mismatches 46; Indels 37; Gaps 5;
QY 12 KRFVSEAEI-----DEKRRROEWEKVRKPED-----PEECPEVYDPRSLYER 57
Db 1004 EKRYREBELQOEEBQLRERERKRROGWEROYRKKDLOQEEBQLRERERKRROER 1063
QY 58 -----LOEKD-----RKQOYEVEORFKFMVAGLDEDETNFIDEVSRQOEL 99
Db 1064 EROYREBELQOEEBQLGEBRETRRROELRORYRKEBELQOEEBQLRERERKRROER 1123
QY 100 IEKORREBELKE-----LKEYRNNLKKYISQENKKEVE 133
Db 1124 EROCREBELQOEEBQLRERERKRROELRORYREBE 1162

RESULT 3
US-08-800-644-94
; Sequence 94, Application US/0800644
; Patent No. 5958752
; GENERAL INFORMATION:
; APPLICANT: Steinert, Peter M.
; APPLICANT: Lee, Seung-Chul
; APPLICANT: Kim, In-Gyu
; APPLICANT: Chung, Soo-Il
; APPLICANT: Park, Sang-Chul
; TITLE OF INVENTION: Trichohyalin and Transglutaminase-3 and
; NUMBER OF SEQUENCES: 117
; TITLE OF INVENTION: Methods of Using Same
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe, Martens, Olson & Bear
; STREET: 620 Newport Center Drive, Sixteenth Floor
; CITY: Newport Beach
; STATE: CA
; COUNTRY: U.S.A.
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/800, 644
; FILING DATE: 14-FEB-1997
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/056, 200
; FILING DATE: 30-APR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Fedrick, Michael F.
; REGISTRATION NUMBER: 36,799
; REFERENCE/DOCKET NUMBER: NIH054.001A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (714) 760-0404
; TELEFAX: (714) 760-9502
; INFORMATION FOR SEQ ID NO: 94:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1898 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-800-644-94

Query Match 10.1%; Score 131.5; DB 2; Length 1898;
Best Local Similarity 25.8%; Pred. No. 0.0039;
Matches 41; Conservative 35; Mismatches 46; Indels 37; Gaps 5;
QY 12 KRFVSEAEI-----DEKRRROEWEKVRKPED-----PEECPEVYDPRSLYER 57
Db 1004 EKRYREBELQOEEBQLRERERKRROGWEROYRKKDLOQEEBQLRERERKRROER 1063
QY 58 -----LOEKD-----RKQOYEVEORFKFMVAGLDEDETNFIDEVSRQOEL 99
Db 1064 EROYREBELQOEEBQLGEBRETRRROELRORYRKEBELQOEEBQLRERERKRROER 1123
QY 100 IEKORREBELKE-----LKEYRNNLKKYISQENKKEVE 133
Db 1124 EROCREBELQOEEBQLRERERKRROELRORYREBE 1162
RESULT 4
US-08-095-737-4
; Sequence 4, Application US/08095737
; Patent No. 5487979
; GENERAL INFORMATION:
; APPLICANT: Difiore, Pier P
; APPLICANT: Fazioli, Francesca

INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 542 amino acids
TYPE: amino acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-701-380-2

Query Match 9.2%; Score 120.5; DB 1; Length 542;
Best Local Similarity 22.4%; Pred. No. 0.0066;
Matches 55; Conservative 33; Mismatches 107; Indels 51; Gaps 9;

QY 25 RKRROEWE-----KVRKP-----EDPECEPEYVDPRSLYERLQEQDRKQOE 68
DB 42 RKRRTAPESCPGSKRKRKAGRRGRPREPSDPQAARAPQTVARFLRDEAKKRD 101
QY 69 YEEQFKKNMVRGLDEDTNFLDEVSRQOELEKOR-----REELKELEKERNN 118
DB 102 PREFTVARAPDADDEEEDEDEDEEAEKKEKILLPPKKPLREKSSADLKERRAK 161
QY 119 LKVGISOENKKEY-EKKLVKPIETKNKFSQAKLAGAVKHSSESGNSVKRLKPDPEP 177
DB 162 -----AQGRGDLGSDPPPKPLRVNKK--EAPAGECTMKRRTKKKSGEADKDPGSGP 213
QY 178 DDKNOEPPSSCKSLGNTSLGSPSHCPSAVVCIGILPGLGAVSGSSDSESSDSECTI--- 234
DB 214 ASARKSPAMFLVGEKS-----PDKKA---LKKKGTPKGARKKEEEDDEEAATYIKN 261
QY 235 -NATGK 239
DB 262 SNQKGK 267

RESULT 9

US-09-032-365A-13
Sequence 13, Application US/09032365A
Patent No. 6114502
GENERAL INFORMATION:
APPLICANT: No. 6114502th, Michael
APPLICANT: Nishina, Patsy
APPLICANT: Nagart, Juergen
APPLICANT: No. 6114502en-Trauth, Konrad
TITLE OF INVENTION: GENE FAMILY ASSOCIATED WITH
NUMBER OF SEQUENCES: 67
CORRESPONDENCE ADDRESS:
ADDRESSEE: Bozicevic & Reed, LLP
STREET: 285 Hamilton Avenue, Suite 200
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94301
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/032,365A
FILING DATE:
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Sherwood, Pamela J
REGISTRATION NUMBER: 36,677
REFERENCE/DOCKET NUMBER: SEQ-2C1IP2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-327-3400
TELEFAX: 650 327-3231

TELEX:
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 542 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-032-365A-13

Query Match 9.2%; Score 120.5; DB 3; Length 542;
Best Local Similarity 22.4%; Pred. No. 0.0066;
Matches 55; Conservative 33; Mismatches 107; Indels 51; Gaps 9;

QY 25 RKRROEWE-----KVRKP-----EDPECEPEYVDPRSLYERLQEQDRKQOE 68
DB 42 RKRRTAPESCPGSKRKRKAGRRGRPREPSDPQAARAPQTVARFLRDEAKKRD 101
QY 69 YEEQFKKNMVRGLDEDTNFLDEVSRQOELEKOR-----REELKELEKERNN 118
DB 102 PREFTVARAPDADDEEEDEDEDEEAEKKEKILLPPKKPLREKSSADLKERRAK 161
QY 119 LKVGISOENKKEY-EKKLVKPIETKNKFSQAKLAGAVKHSSESGNSVKRLKPDPEP 177
DB 162 -----AQGRGDLGSDPPPKPLRVNKK--EAPAGECTMKRRTKKKSGEADKDPGSGP 213
QY 178 DDKNOEPPSSCKSLGNTSLGSPSHCPSAVVCIGILPGLGAVSGSSDSESSDSECTI--- 234
DB 214 ASARKSPAMFLVGEKS-----PDKKA---LKKKGTPKGARKKEEEDDEEAATYIKN 261
QY 235 -NATGK 239
DB 262 SNQKGK 267

RESULT 10

US-08-933-750C-47
Sequence 47, Application US/08933750C
Patent No. 5932442
GENERAL INFORMATION:
APPLICANT: Lal, Preeti
APPLICANT: Hillman, Jennifer L.
APPLICANT: Bandman, Olga
APPLICANT: Shah, Puri V
APPLICANT: Au-Young, Janice
APPLICANT: Yue, Henry
APPLICANT: Guegler, Karl J.
APPLICANT: Corley, Neil C.
TITLE OF INVENTION: HUMAN REGULATORY MOLECULES
NUMBER OF SEQUENCES: 98
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/933,750C
FILING DATE: September 23, 1997
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749

TELEX: 904136
; INFORMATION FOR SEQ ID NO: 24:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 285 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
US-08-149-809-24

Query Match 9.1%; Score 119; DB 1; Length 285;
Best Local Similarity 24.5%; Pred. No. 0.0037;
Matches 53; Conservative 41; Mismatches 80; Indels 42; Gaps 9;

QY 7 GNLIKKRFSEAE-----LDERKKRROEEMK-----VRRKPEDPCPEFYDPRS 53
DB 69 GKLEKKKALONAESEVALNRRQLLEEDLESEERLGSATKLSASQADESERARK 128
QY 54 LYERLQPKDRKQOEYEEQFKNMVGRGLDEDTNPLDEVSROELIEK--QRRRE---- 107
DB 129 ILEBRALADEBRDALENQK---EARFLAEADKKYDEVARKLAWAEADLERAEBAEQ 185
QY 108 -----EL-KELEYNNLKKVIGISQ--NKKEYEKLTVKPIETKKNFSQAKLLAGAVK 158
DB 186 GENKIVLEELRYGNNLKSLEYSEKSNQREVEYKNQIKTLNTRLEAE----- 237
QY 159 HKSESGNSVYKRLKDPD--PDDKNOEPSSCKSLGN 192
DB 238 -RAEFARSVOKLOKEVDRLDLELEKERYKIDGD 272

RESULT 13

PCT-US93-07261-11
; Sequence 11, Application PC/TUS9307261
; GENERAL INFORMATION:
; TITLE OF INVENTION: PFEPM3 MALARIA ANTIGEN, ANALOGS, ANTIBODIES AND USES THEREOF
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: John H. C. Blasdale
; STREET: One Giralda Farms
; CITY: Madison
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07940-1000
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Macintosh 6.0.5
; SOFTWARE: Microsoft Word 5.1a
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/07261
; FILING DATE: 19930805
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/927,531
; FILING DATE: 07-AUG-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Blasdale, John H. C.
; REGISTRATION NUMBER: 31,895
; REFERENCE/DOCKET NUMBER: DX0288K
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-822-7398
; TELEFAX: 201-822-7039
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1588 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-US93-07261-11

Query Match 9.0%; Score 118; DB 4; Length 1588;
Best Local Similarity 24.7%; Pred. No. 0.048;
Matches 58; Conservative 40; Mismatches 67; Indels 70; Gaps 13;

QY 5 DDGNLIKK-----RFVSEALDER-----RKRROEEM-----EKVRKPED 40
DB 105 DDKKIYFKKPESEYKDLKELKEKEFIKQHLADYERKKRRNNILRSIRROKLEIEQ 164
QY 41 PE-----ECPEEYVDPRLYERLQ-----EOKDRKQOEYEEQFKNMVGRGLDED 85
DB 165 LEKLNQLESAINELKERRASRRPMVKMGKMDVEDEWIKKYDDEQAKN-----GTKDE 221
QY 86 ETNPLDEVSRQOELIEKQ---RRBEELKELEY-----RNNLKKVIGISQENKKEYEKK 135
DB 222 EIK--DKGDYEIEVEIETKFGMBENALGELDEVEERYEKKRYLYKEDG--EGDLKDYEEK 277
QY 136 LTVKPIETKFKFSQAKLLAGAVKHKSESGNSVYKRLKPD-----PEPDDKNQ 182
DB 278 LEETGEGFERKEFTTRL--VRRKRK--EOKKLKEDEKKLIAAEPPDERK 326

RESULT 14

PCT-US93-07261-16
; Sequence 16, Application PC/TUS9307261
; GENERAL INFORMATION:
; TITLE OF INVENTION: PFEPM3 MALARIA ANTIGEN, ANALOGS, ANTIBODIES AND USES THEREOF
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: John H. C. Blasdale
; STREET: One Giralda Farms
; CITY: Madison
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07940-1000
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Macintosh 6.0.5
; SOFTWARE: Microsoft Word 5.1a
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/07261
; FILING DATE: 19930805
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/927,531
; FILING DATE: 07-AUG-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Blasdale, John H. C.
; REGISTRATION NUMBER: 31,895
; REFERENCE/DOCKET NUMBER: DX0288K
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-822-7398
; TELEFAX: 201-822-7039
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1663 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; ORGANISM: Plasmodium falciparum
; STRAIN: Malayan Camp
PCT-US93-07261-16

Query Match 9.0%; Score 118; DB 4; Length 1663;
Best Local Similarity 24.7%; Pred. No. 0.052;
Matches 58; Conservative 40; Mismatches 67; Indels 70; Gaps 13;

QY 5 DDGNLIKK-----RFVSEALDER-----RKRROEEM-----EKVRKPED 40
DB 105 DDKKIYFKKPESEYKDLKELKEKEFIKQHLADYERKKRRNNILRSIRROKLEIEQ 164
QY 41 PE-----ECPEEYVDPRLYERLQ-----EOKDRKQOEYEEQFKNMVGRGLDED 85
DB 165 LEKLNQLESAINELKERRASRRPMVKMGKMDVEDEWIKKYDDEQAKN-----GTKDE 221

QY 86 ETNFDLVSROQLIEKQ---RREELKELKEY-----RNNLKVGISQENKKEVEKK 135
DB 222 EIK--DKGGYEIEYETLKYGKRENALGELDEYERKRYLLKEDG--EGDLKDVEEK 277
QY 136 LTVKPIETKRNKSQAAILAGAVKHKSSGNSVVKRLKP-----PEPDCKNO 182
DB 278 LEETGYGREKFPPTTRIL--VKRRNK--EOKKLEKDEKRLIAEPPDDEKK 326

RESULT 15

US-08-741-134-6
; Sequence 6, Application US/08741134
; Patent No. 5861498
; GENERAL INFORMATION:
; APPLICANT: Litwack, Gerald
; APPLICANT: Alnemri, Emad S.
; APPLICANT: Fernandes-Alnemri, Teresa
; TITLE OF INVENTION: IMMUNOPHILIN FKBP46 AND COMPOSITIONS FOR MAKING
; TITLE OF INVENTION: AND
; TITLE OF INVENTION: METHODS OF USING THE SAME
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 5861498rls
; STREET: One Liberty Place - 46th floor
; CITY: Philadelphia
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: Windows 3.11
; SOFTWARE: Wordperfect for Windows 6.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/741,134
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/007,163
; FILING DATE: 01-NOV-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Deluca, Mark
; REGISTRATION NUMBER: 33,229
; REFERENCE/DOCKET NUMBER: TJU-2090
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-568-3100
; TELEFAX: 215-568-3439
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 411 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-741-134-6

Query Match 8.9%; Score 116; DB 2; Length 411;

Best Local Similarity 23.0%; Pred. No. 0.011; Mismatches 79; Indels 76; Gaps 9;

QY 3 GDDGNLIIKKRVSEALDERRRKROEWEKVRK-----PEDPECEPE--- 47
DB 68 GGD-----FDEDEIDESSSEEE-EXTOKKKSKGKKAESSEDEDEDEDE 116
QY 48 -----YDPRSLYERLOEQKDKQOYE-----EOKFKNNVRCID 83
DB 117 FOESVLLTLSPQAQYQOSIDLITPPEEYQFIVTGSYAISLSGNYVKHFDTPMGVEGED 176
QY 84 EDE-----TNFDLVSROQLI-----EKORREELKELKEYRNNLKKVGISQ 126
DB 177 EDEADIDYDSEYDITPDEDEITIGDDMDLDEDEEEVRIEYQDEDEEDNDGEE---EQ 233

QY 127 ENKKEVEKKLTVKPIETKRNKSQAAILAGAVKHKSSGNSVVKRLKPPEPDCKNOEPSS 186
DB 234 EEEEEEQKEEYKPPPKSKKKEKR-----KHEKEEERKAKVKVYFKKDLBEGPTK 287
QY 187 CKS 189
DB 288 PKs 290

Search completed: February 16, 2001, 17:06:52
Job time: 105 sec

~~BEST AVAILABLE COPY~~

THIS PAGE BLANK (UST)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2001, 17:05:07 ; Search time 25.95 Seconds
(without alignments)
1147.237 Million cell updates/sec

Title: US-09-602-597-2
Perfect score: 1306
Sequence: 1 MDGGDDGILIKRRFVSEAE.....NATGKIVSIFRTNPLEAP 254

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 374700 seqs, 117207915 residues
Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Maximum Match 0%
Listing first 45 summaries

Database :
1: SP archaea:*
2: SP bacteria:*
3: SP fungi:*
4: SP human:*
5: SP_invertebrate:*
6: SP_mammal:*
7: SP_mhc:*
8: SP_organelle:*
9: SP_phage:*
10: SP_plant:*
11: SP_rodent:*
12: SP_virus:*
13: SP_vertebrate:*
14: SP_unclassified:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	422.5	32.4	249	5	09V843
2	302.5	23.2	219	5	017594
3	169.5	13.0	241	10	09M180
4	160.5	12.3	2274	5	09VYU0
5	151.5	11.6	3476	5	09NM17
6	151.5	11.6	5533	5	09U6C3
7	151.5	11.6	5554	5	09NMN1
8	151	11.6	651	4	09UPZ5
9	150.5	11.5	693	5	09VNG4
10	149.5	11.4	312	5	09V843
11	149	11.4	1087	5	09V843
12	148	11.3	314	5	015987
13	147	11.3	706	11	09QX56
14	146	11.2	271	5	044077
15	146	11.2	293	5	015988
16	146	11.2	451	2	050870
17	146	11.2	460	2	09ZIU2
18	145.5	11.1	699	10	09ZU69
19	145.5	11.1	2701	4	09Y520

20	144.5	11.1	513	10	09LW95	091W95 nicotiana t
21	143	10.9	661	11	070205	070205 rattus norv
22	143	10.9	1165	4	095819	095819 homo sapien
23	143	10.9	2081	10	09LH98	091H98 arabidopsis
24	142	10.9	772	13	091013	091013 gallus gall
25	142	10.9	1661	5	006166	006166 plasmodium
26	141.5	10.8	384	5	09XZ71	09XZ71 periplaneta
27	141.5	10.8	466	5	09J312	09J312 caenorhabd
28	141	10.8	1027	4	09NS77	09NS77 homo sapien
29	141	10.8	1257	4	09S033	09S033 homo sapien
30	140.5	10.8	1175	4	075172	075172 homo sapien
31	140	10.7	380	5	09XZ72	09XZ72 libellula p
32	139.5	10.7	1183	11	089040	089040 rattus norv
33	139	10.6	1233	11	097820	097820 mus musculu
34	138.5	10.6	335	5	076719	076719 caenorhabd
35	138	10.6	385	5	017909	017909 caenorhabd
36	137	10.5	1359	5	002061	002061 caenorhabd
37	137	10.5	1359	5	0907E0	0907E0 caenorhabd
38	136.5	10.5	673	5	09U0N1	09U0N1 plasmodium
39	136	10.4	660	11	090X55	090X55 mus musculu
40	135	10.3	729	3	09P3P1	09P3P1 neosporea
41	135	10.3	777	5	045373	045373 caenorhabd
42	135	10.3	919	4	09NQ57	09NQ57 homo sapien
43	134.5	10.3	1010	5	09VB01	09VB01 drosophila
44	134	10.3	554	5	09VPS3	09VPS3 drosophila
45	133.5	10.2	735	5	09U0D5	09U0D5 tetrahymena

ALIGNMENTS

RESULT	1	PRELIMINARY:	PRT:	249 AA.
ID	09V843	AC	09V843	
DT	01-MAY-2000 (TREMBLrel. 13, Created)	DT	01-MAY-2000 (TREMBLrel. 13, Last sequence update)	
DT	01-MAY-2000 (TREMBLrel. 13, Last annotation update)	DT	01-MAY-2000 (TREMBLrel. 13, Last annotation update)	
DE	CG14480 PROTEIN.	DE	CG14480 PROTEIN.	
GN	CG14480.	GN	CG14480.	
OS	Drosophila melanogaster (Fruit fly).	OS	Drosophila melanogaster (Fruit fly).	
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;	OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;	
OC	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;	OC	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;	
OC	Ephydroidea; Drosophilidae; Drosophila.	OC	Ephydroidea; Drosophilidae; Drosophila.	
OX	NCBI_TaxID=7227;	OX	NCBI_TaxID=7227;	
RP	[1]	RP	[1]	
RN	SEQUENCE FROM N.A.	RN	SEQUENCE FROM N.A.	
RC	STRAIN=BERKELEY;	RC	STRAIN=BERKELEY;	
RX	MEDLINE=20196006; PubMed=10731132;	RX	MEDLINE=20196006; PubMed=10731132;	
RA	Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,	RA	Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,	
RA	Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,	RA	Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,	
RA	George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,	RA	George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,	
RA	Sutton G.G., Mortman J.R., Yandell M.D., Zhang Q., Chen L.X.,	RA	Sutton G.G., Mortman J.R., Yandell M.D., Zhang Q., Chen L.X.,	
RA	Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,	RA	Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,	
RA	Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,	RA	Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,	
RA	Abriil J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,	RA	Abriil J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,	
RA	Bailey R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,	RA	Bailey R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,	
RA	Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,	RA	Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,	
RA	Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,	RA	Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,	
RA	Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,	RA	Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,	
RA	Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,	RA	Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,	
RA	de Pablos B., Delcher A., Deng Z., Meyers A.D., Dew I., Dietz S.M.,	RA	de Pablos B., Delcher A., Deng Z., Meyers A.D., Dew I., Dietz S.M.,	
RA	Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,	RA	Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,	
RA	Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,	RA	Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,	
RA	Foster C., Gabrielian A.E., Gary N.S., Gelbart W.M., Glasser K.,	RA	Foster C., Gabrielian A.E., Gary N.S., Gelbart W.M., Glasser K.,	
RA	Glocke A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,	RA	Glocke A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,	
RA	Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck C.,	RA	Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck C.,	
RA	Hoslin D., Houston K.A., Howland T.J., Wei M.-H., Ibegyan C.,	RA	Hoslin D., Houston K.A., Howland T.J., Wei M.-H., Ibegyan C.,	
RA	Jatali M., Kalush F., Karpen G.H., Ke Z., Kenison J.A., Ketchum K.A.,	RA	Jatali M., Kalush F., Karpen G.H., Ke Z., Kenison J.A., Ketchum K.A.,	
RA	Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,	RA	Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,	
RA	Lasero P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,	RA	Lasero P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,	
RA	Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,	RA	Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,	
RA	Merkulov G., Mishaeva N.V., Modyary C., Morris J., Moshrefi A.,	RA	Merkulov G., Mishaeva N.V., Modyary C., Morris J., Moshrefi A.,	

RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon R., Nusskern D.R., Paley J.M.,
 RA Palazzolo M., Piaton G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shie B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venier E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,"
 RT "The genome sequence of *Drosophila melanogaster*.";
 RL Science 287:2185-2195(2000).
 DR EMBL: AE003802; AAF57835.1; -
 DR FLYBASE: FBgn0034242; CG14480.
 SO SEQUENCE 249 AA; 28376 MW; 070FA9AE2764849E CRC64;

Query Match 32.4%; Score 422.5; DB 5; Length 249;
 Best Local Similarity 41.8%; Pred. No. 4.2e-18;
 Matches 94; Conservative 44; Mismatches 76; Indels 11; Gaps 5;

QY 15 FVSEELDERRRRROEEMKVKRPDPPECEPEYVDPRSLYERLOEQDRKQOEYEDFK 74
 DB 5 FVTEAEAEQROREWEVERVQPEDPLEPEPYDGRSLYERLQNDKDMEEFAHK 64
 QY 75 FKNMVRGDEDETNFLDEYSRQOELIEKORREELKELEYNNKKVGIQSENNK-EVE 133
 DB 65 LKNLRIGLDEDESVFLDYDAHKIHAERQOMDELELDFNNRREKLOEESVDDKLOE 124
 QY 134 KKLIVKPI-ETKNNKPSQAKLLAGAVKHKSSSGNSVRLK-----PPPEDDKNQEPSS 186
 DB 125 LKTTAKSAGASVGRSTQGLLGGIKRKNGELPTTSVAKITENEVEQATNEATKRPAD 184
 QY 187 CKSLGNTSLSGPSIHCPAANVCIGILPGLGAVSGSSDSESSDSE 231
 DB 185 -KTTWTLITNKK-YNQAGALKCIAILPGISTESSDSEASTDEE 226

RESULT 2
 017594 PRELIMINARY; PRT; 219 AA.
 ID 017594;
 AC 017594;
 DT 01-JAN-1998 (TREMBlrel. 05, Created)
 DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
 DT 01-JAN-1999 (TREMBlrel. 09, Last annotation update)
 DE C25A1.1 PROTEIN.
 GN C25A1.1.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabdilitida; Rhabdilitidae;
 OC Rhabdilitidae; Peioderitidae; Caenorhabditis.
 OC NCBI_Taxid=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Mortimore B.;
 RL Submitted (OCT-1996) to the EMBL/Genbank/DBJ databases.
 RP MEDLINE FROM N.A.
 RX MEDLINE=94150716; PubMed=7906398;
 RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
 RA Bonfield J., Burton J., Connell M., Copey T., Cooper J., Coulson A.,
 RA Brockton M., Dear S., Du Z., Durbin R., Faveille A., Fulton L.,
 RA Gardner A., Green P., Hawkins J., Hillier L., Jier M., Johnston L.,
 RA Jones M., Kershaw J., Kirsten T., Laister N., Latreille P.,
 RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
 RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Showkeen R.,
 RA Smalton N., Smith A., Sonhammer E., Staden R., Sulston J.,
 RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
 RA Watson A., Weinstock L., Wilkinson-Sprat J., Woldman P.;
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of *C. elegans*.";
 RL Nature 368:32-38(1994).

DR EMBL: Z81038; CAB02762.1; -
 SQ SEQUENCE 219 AA; 24353 MW; 9DD37690588FDE85 CRC64;

Query Match 23.2%; Score 302.5; DB 5; Length 219;
 Best Local Similarity 36.4%; Pred. No. 4.4e-11;
 Matches 82; Conservative 36; Mismatches 70; Indels 37; Gaps 8;

QY 15 FVSEELDERRRRROEEMKVKRPDPPECEPEYVDPRSLYERLOEQDRKQOEYEDFK 74
 DB 5 FVTEAEAEQROREWEVERVQPEDPLEPEPYDGRSLYERLQNDKDMEEFAHK 64
 QY 75 FKNMVRGDEDETNFLDEYSRQOELIEKORREELKELEYNNKKVGIQSENNK-EVE 132
 DB 65 LKNMVRGDEDESVFLDYDAHKIHAERQOMDELELDFNNRREKLOEESVDDKLOE 117
 QY 133 EKKLIVKPIETK-----NFKSQAKLLAGAVKHKSSSGNSVRLKRPDPPEDDKNQEPSSCK 188
 DB 118 SSRFLIKPSTSKVLGPSPKQAAFLSTAIKRKSTST-----EKKOE-- 159
 QY 169 SLGNTSLSGPSIHCPAANVCIGILPGLGAVSGSSDSESSDSE 231
 DB 160 DVSSKSKPE-----PVIKQIGALQALCEYDPSSESDESDASSDE 200

RESULT 3
 09M1R0 PRELIMINARY; PRT; 241 AA.
 ID 09M1R0;
 AC 09M1R0;
 DT 01-OCT-2000 (TREMBlrel. 15, Created)
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
 DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
 DE HYPOTHETICAL 27.2 KDA PROTEIN.
 GN T17J13.100.
 OS Arabidopsis thaliana (mouse-ear cress).
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
 OC Brassicales; Brassicaceae; Arabidopsis.
 OC NCBI_Taxid=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Rieger M., Mueller-Auer S., Zipp M., Schaefer M., Mewes H.W.,
 RA Lemcke K., Mayer K.F.X., Quetler F., Salanoubat M.;
 RL Submitted (FEB-2000) to the EMBL/Genbank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA EU Arabidopsis sequencing project;
 RL Submitted (FEB-2000) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AL138651; CAB71874.1; -
 KW Hypothetical protein.
 SO SEQUENCE 241 AA; 27168 MW; 350F997029307F14 CRC64;

Query Match 13.0%; Score 169.5; DB 10; Length 241;
 Best Local Similarity 23.3%; Pred. No. 0.0034;
 Matches 63; Conservative 30; Mismatches 78; Indels 99; Gaps 8;

QY 15 FVSEELDERRRRROEEMKVKRPDPPECEPEYVDPRSLYERLOEQDRKQOEYEDFK 74
 DB 14 FVSEQDSESKRERGERVEDGTFORD-----RALTVELIKENDKDAEFNEKFK 62
 QY 75 FKNMV-----RGDEDETNFLDEYSRQOELIEKORREELKELEYNNKKVGIQSENNK-EVE 108
 DB 63 HSESLSTVSSIDSVSCDQFLCQDLCGPPKALDEDETEFLDKL----- 107
 QY 109 LKELEKYYNNKKVGIQSENNKVEYKUTVPIETKNNKPSQAKLLAGAVKHKSSSGNSV 168
 DB 108 -----EMSKREYERQLANEDEDQLRNFOQAAVARSAIILHEPKEA-- 147
 QY 169 KRLKPPDPDDKNQEPSSCKSLGNTSLSGPSIHCPAANVCIGILPGLGAVSGSSDSESSDSE 214
 DB 148 -LPPAPVTKQKP-----IGKRN--PATRPFKALIKVKKPPKAKATEKEKEKEIPG 196

```

Db 1477 LREKIKEKEKE--KLREEREKMKEREKEKIKERVEKIKEREEREKLTKEKEEKL 153
      :: : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

GN SPEN.
OS *Drosophila melanogaster* (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Eukaryota; Metazoa; Echinodermata; Chordata; Vertebrata; Mammalia

RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.H.C., Blazey R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Mikos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Plankkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Bakendale J., Bayraktaroglu I., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Bottier P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahle C., Daveport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fliselman W.,
 RA Foster C., Gabriellan A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mates J., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Mishina N.V., Moberly C., Morris J., Moshireli A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M., Wang X.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter G., Wang A.H., Wang X.,
 RA Wang Z.-Y., Massarman D.A., Weinstein G.M., Weissenbach J.,
 RA Williams S.M., Woodage T., Morley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs S.A., Myers F.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of *Drosophila melanogaster*.";
 RL Science 287:2185-2195(2000).
 CC -1- ALTERNATIVE PRODUCTS: TWO ISOFORMS: LONG ISOFORM (SHOWN HERE) AND
 CC SHORT ISOFORM: ARE PRODUCED BY ALTERNATIVE SPLICING.
 DR EMBL: AEO03602; AAF51972.1; -;
 DR EMBL: AEO03602; AAF51973.1; -;
 DR FLYBASE: FBgn0037362; CG2179.
 KW Alternative splicing; Hypothetical protein.
 FT VARSPLIC 1 135 MISSING (IN SHORT ISOFORM).
 SQ SEQUENCE 693 AA; 80212 MW; F65BA9601F086EDD CRC64;

Query Match 11.5%; Score 150.5; DB 5; Length 693;
 Best Local Similarity 23.1%; Pred. No. 0.13; Mismatches 83; Indels 61; Gaps 8;
 Matches 58; Conservative 49;

QY 11 IKRREYSEALDERRRKROEEMEKVKRPDEPCEPEVYDPSLYERLQDQKROQOE 70
 DB 277 IRRKFFAKA-----KAEEDEERKLNKQOE-----ELDRERQKOEELKAEKAEKQERE 325
 QY 71 EOKPFKNMVGLEDETFNLFDEV--SRQOELEKORREELKELK-----EYNNNKKYV 123
 DB 326 ERKKEKMKKIOERGVEISQKIRIEKRLIMQORLESITRLLEKLFERTQLQDRKNRR 395
 QY 124 ISOENKKEVEKKLTVPRIETKKNFSQAKL-----TAGAVKHKSSSGNS 167
 DB 386 VKQDDDKDIORGLVE-----KYKAATEKLVCDOKRIVQETKSMTPPLGLLKSCKSKSA-- 439
 QY 168 VKRLKPDPEDDKNOEPSSCKSLGNTSLSGPSIHCS-----AAVCIGILPG 214
 DB 440 -----PVDASSEDEAASKRHRKMGNGVTESAPPGVNSALNPLNVAAMAAGAVPG 491
 QY 215 IGAYSGSDSE 225
 DB 492 PGGYSAKAVASE 502

RESULT 10

P91570
 ID P91570 PRELIMINARY; PRT; 312 AA.
 AC P91570;
 DT 01-MAY-1997 (TREMBlrel. 03, Created)
 DT 01-MAY-1997 (TREMBlrel. 03, Last sequence update)
 DT 01-NOV-1998 (TREMBlrel. 08, Last annotation update)
 DE COSMID ZK354.
 GN ZK354.3
 OS *Caenorhabditis elegans*.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
 OC Rhabditidae; Peloderinae; *Caenorhabditis*.
 OX NCBI_Taxid=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RX MEDLINE=94150718; PubMed=7906398;
 RA Wilson R., Almscough R., Anderson K., Baynes C., Berks M.,
 RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
 RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
 RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
 RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
 RA Lightning J., Lloyd C., Mcmuray A., Mortimore B., O'Callaghan M.,
 RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showmken R.,
 RA Smaldon N., Smith A., Sonhammer E., Staden R., Sulston J.,
 RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
 RA Watson A., Weinstein L., Wilkinson-Sproat J., Wohlman P.,
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of *C. elegans*.";
 RL Nature 368:32-38(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA Johnson D., Kamsley P., Bradshaw H.;
 RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA Waterston R.;
 RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL: U08172; AAB42258.1; -;
 SQ SEQUENCE 312 AA; 36737 MW; 958C146E0021E42E CRC64;

Query Match 11.4%; Score 149.5; DB 5; Length 312;
 Best Local Similarity 24.9%; Pred. No. 0.066; Mismatches 78; Indels 19; Gaps 5;
 Matches 46; Conservative 42;

QY 20 ELDERRRKROEEMEKVKRPDEPCEPEVYDPS-----LYERLQDQKROQ 66
 DB 68 EKKEEKEEKEEKEEKKADDEKKTTEKDKSKSTEEDKTSYKKTQETKSEKDKD 127
 QY 67 QEYEOFFKFNWVGLEDETFNLFDEV--SRQOELEKORREELKELKRYNNKKVGI 124
 DB 128 ERKDEKKEEENKESKDEEKK--DEVKDKDEKDEKMDKKPKGEKKEEKKEMKEEK 185
 QY 125 SQENKKEVEKK-LTVPRIETKKNFSQAKLAGAVKHKSSSGSVRLKADPEDDKNOE 183
 DB 186 KEKKEEPPKNDAPKKEGTEGVKDALVENDPIVTEMSDR-DEKKEKKDKDKDEKDE 244
 QY 184 PSSCK 188
 DB 245 KKEAK 249

RESULT 11
 ID 096923 PRELIMINARY; PRT; 1087 AA.
 AC 096923;
 DT 01-MAY-1999 (TREMBlrel. 10, Created)
 DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
 DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
 DE GELSOLIN-RELATED PROTEIN.
 GN GRP120.

OS Dictyostelium discoideum (slime mold).
 OC Eukaryota; Dictyostelida; Dictyostelium.
 OX NCBI_TaxID=44689;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=AX2;
 RA Stocker S., Marriott G.;
 RT "GRP120 from Dictyostelium discoideum represents a new type of a
 RL gelsolin-related protein."
 RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL: U95159; AAD00774.1; -
 DR INTERPRO: IPR001974; -
 DR PRAM: PF00626; Gelsolin; 1.
 DR PRINTS: PR00597; GELSOLIN.
 SQ SEQUENCE 1087 AA; 124777 MW; 8C431F5A0B51E677 CRC64;

Query Match 11.4%; Score 149; DB 5; Length 1087;
 Best Local Similarity 24.1%; Pred. No. 0.25; Mismatches 95; Indels 32; Gaps 9;
 Matches 59; Conservative 59; Mismatches 95; Indels 32; Gaps 9;

QY 11 IKKRFVSEALDERRRKROEEMKVRKPED---PEECPEEYDPRSLYERLQOKDRKQ 67
 Db 717 LAELQKEKEDLEKQQQEQEQEQKNNKIYEEVKEEYKE-EDVKEEVEEVEE 775
 QY 68 EYERQFK-----FKNNVRLGDEDETFLDEVSRQOELIEKORREELK-ELK--EYRNN 118
 Db 776 VKEEVEKVAKEETKEEKEVNDKATEVKNQVEEVEEVEEVEEVEEVEEYKGE 835
 QY 119 LKKYIGIQENKKEVEKLYKPT--ETKNNFSQAK--LLAGAVKHKSSGNSVRLKPD 174
 Db 836 AKEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEE 895
 QY 175 PEPDKNQEPSSCSLGTSLSPSICHPSAVCGILPGAVSGSSDSSDSEGTI 234
 Db 896 EKVAEENETVAEENEG-ILVSPSEKVEDA-----NSSSTISSPENEGSV 940
 QY 235 NATGK 239
 Db 941 SVKDK 945

RESULT 12
 ID 015987 PRELIMINARY; PRT; 314 AA.
 AC 015987;
 DT 01-JAN-1998 (TREMBlrel. 05, Created)
 DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
 DT 01-MAY-2000 (TREMBlrel. 13, Last annotation update)
 DE TROPONIN I.
 OS Pectinopecten yessoensis (Ezo giant scallop) (Yesso scallop).
 OC Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorpha; Pectinoida;
 OC Pectinoidea; Pectinidae; Pectinopecten.
 OX NCBI_TaxID=6573;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Hotta A.L.;
 RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AB008005; BAA22852.1; -
 DR INTERPRO: IPR001978; -
 DR PRAM: PF00992; Tropoin; 1.
 SQ SEQUENCE 314 AA; 37009 MW; 2FB61D644BCAED8 CRC64;

Query Match 11.3%; Score 148; DB 5; Length 314;
 Best Local Similarity 28.1%; Pred. No. 0.081; Mismatches 72; Indels 28; Gaps 4;
 Matches 50; Conservative 28; Mismatches 72; Indels 28; Gaps 4;

QY 3 GCDGNTLIKKRFVSEALDERRRKROEEMKVRKPEDPEECPEEYDPRSLYERLQOK 62
 Db 58 GNDSDYSAAAEPRAYDADNRKROQEEBEAARAAEE-----YKROEEL 105
 QY 63 DRKQOEYEQFKFNWVRLGDEDETFLDEVSRQOELIEKORREELKELKEYRNNLKV 122

Db 106 RRORREERORREERORROQEEERLRLER-----EEOEREEARMAEQQKKKK 157
 QY 123 ---GISQENKKEVEKKLVKPIETKNNKSOAKLLA-----GAVKHKSSGNSVRLK 172
 Db 158 GLGSLSPKKMKLKLWQKAEDLKNAEAKKEKYINDLVKSTDSKDVAAALQ 215

RESULT 13
 ID 090XS6 PRELIMINARY; PRT; 706 AA.
 AC 090XS6;
 DT 01-MAY-2000 (TREMBlrel. 13, Created)
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
 DT 01-JUN-2000 (TREMBlrel. 14, Last annotation update)
 DE DREBRIN A.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SV129;
 RA Keon B.H., Jedrzejewski P.T., Paul D.L., Goodenough D.A.;
 RT "Isoform specific expression of the neuronal F-actin binding protein,
 RT drebrin, in specialized cells of stomach and kidney epithelia."
 RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF187147; AAF25189.1; -
 DR INTERPRO: IPR002108; -
 DR PRAM: PF00241; Cofilin; 1.
 SQ SEQUENCE 706 AA; 77259 MW; C9E04FEE651D87FB CRC64;

Query Match 11.3%; Score 147; DB 11; Length 706;
 Best Local Similarity 25.5%; Pred. No. 0.21; Mismatches 79; Indels 50; Gaps 11;
 Matches 61; Conservative 49; Mismatches 79; Indels 50; Gaps 11;

QY 17 SEALDERRRKROEEMKVRKPEDPEECPEEYDPRSLYERLQOKDRKQOEYEQFK 76
 Db 166 TDAVEMKRIHQFPWQAKK---EELRKEEERKKLLDALRFEQEMERQEQ----- 218
 QY 77 NMVRLGDEDETFLDEVSRQOELIEKORREELKELKEYRNNLKVGISQENKKEVEKKL 136
 Db 219 -----EERERRRRE---REQO-IEHRRKQOOLEAEKAKRLKQSLFGQORDEE--- 265
 QY 137 TVKPIETKNNKSOAKLLAGAVKHKSSGNSVRLKPD-PEPDKNQEPSSCSLGTSL 195
 Db 266 -----ESQMKSESEV-----EEAAIIRQDPNPFREFRQDERVASAGGSCDA 310
 QY 196 SGPS-ITCPSAVCGILPGIGAY-SGSSDSSDSESG-----TINATGKIVSSI 244
 Db 311 PAPAFNHRPGRPYC---PRKASDSGSSSSSSSSSSSPPTPPYITCHRTPLMSSL 365

RESULT 14
 ID 044077 PRELIMINARY; PRT; 271 AA.
 AC 044077;
 DT 01-JUN-1998 (TREMBlrel. 06, Created)
 DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)
 DT 01-MAY-2000 (TREMBlrel. 13, Last annotation update)
 DE TROPONIN I (FRAGMENT).
 OS Chlamys nipponensis akazara (Akazara scallop) (Japanese scallop).
 OC Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorpha; Pectinoida;
 OC Pectinoidea; Pectinidae; Chlamys.
 OX NCBI_TaxID=6571;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Nishita K., Ojima T., Soejima T.;
 RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AB009368; BAA23775.1; -
 DR INTERPRO: IPR001978; -
 DR PRAM: PF00992; Tropoin; 1.

THIS PAGE BLANK